The Wiley Handbook of Human Computer Interaction

The Wiley Handbook of Human Computer Interaction

Volume 1

Edited by

Kent L. Norman and Jurek Kirakowski

WILEY Blackwell

This edition first published 2018

Registered Office(s)
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Office
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloging-in-Publication Data Is Available

9781118976135 – Hardback
9781118977262 – e-PDF
9781118977279 – e-Pub

Cover Design: Wiley
Cover Image: © Andrea Danti/Shutterstock

Set in 10/12pt Galliard by SPi Global, Pondicherry, India

10 9 8 7 6 5 4 3 2 1

Contents

Notes on Contributors

Marc Abrams serves as Harmonia's president and chief technical officer. He provides technical and business leadership to the company and manages all its technical projects. In the past, Dr. Abrams has been with the former U.S. Army Concepts Analysis Agency, a postdoc in the Distributed Operating Systems group in Stanford's Computer Science Department, and a visiting scientist in the network protocol group at IBM's Zurich Research Laboratory in Switzerland. He has been the principal investigator for over $30 million in research and development projects with the Air Force, the Army, DARPA, DHS, DOE, DOT, NASA, the Navy, NIH, NSF, MDA, ONR, OSD, and various companies including General Dynamics, IBM, Northrop Grumman, Raytheon, and Leidos. He received his PhD from the University of Maryland at College Park in computer science. Before Harmonia, Dr. Abrams was a tenured associate professor at Virginia Tech, where his research on human-computer interfaces (HCI) led to the creation of User Interface Markup Language (UIML) and later the co-founding of Harmonia. UIML forms the basis for Harmonia's LiquidApps® product. At Virginia Tech, he also co-founded the Center for Human Computer Interaction, and worked with the HCI faculty in fields ranging from cognitive psychology to human factors on scenario-driven HCI design.

Nigel Bevan is an independent user experience (UX) consultant with wide industrial and research experience. He has been the editor of several international standards including both the original and revised versions of ISO 9241-11 (usability), 9241-210 (human-centered design processes), 25010 and 25022 (software quality model and measures), 20282-2 (usability test method), and 25063 (context of use). He has authored over 80 publications and was a member of the U.S. National Academy of Science Committee on Human-System Design Support for Changing Technology.

Frédéric Bevilacqua is the head of the Sound Music Movement Interaction team at Institute for Research and Coordination in Acoustics/Music (IRCAM) in Paris, which is part of the joint research lab Science and Technology for Music and Sound (IRCAM—CNRS—Université Pierre et Marie Curie). He received his PhD in Biomedical Optics from EPFL (Swiss Federal Institute of Technology in Lausanne), in 1998. His research concerns the modeling and the design of interaction between movement and sound, and the development of gesture-based interactive systems. With his team, he developed several digital musical instruments such as the augmented violin and the modular musical objects (First Prize of the Guthman Musical Instrument Competition), and developed several systems to synchronize motion to sound, such as the gesture follower. He coauthored more than 120 scientific publications and coauthored five patents. He was keynote or invited speaker at several international conferences such as the ACM TEI'13. As the coordinator of the "Interlude Project" he received the ANR Digital Technology Prize (Societal Impact) in 2013.

Nadia Bianchi-Berthouze is a full professor in affective computing and interaction at the Interaction Centre of the University College London (UCL). She received her PhD in computer science for biomedicine from the University of the Studies of Milan, Italy. Her research focuses on designing technology that can sense the affective state of its users and use that information to tailor the interaction process. She has pioneered the field of affective computing and for more than a decade she has investigated body movement, and more recently touch behavior, as a means to recognize and measure the quality of the user experience in full-body computer games, physical rehabilitation, and textile design. She also studies how full-body technology and body sensory feedback can be used to modulate people's perception of themselves and of their capabilities to improve self-efficacy and copying capabilities. She has published more than 170 papers in affective computing, HCI, and pattern recognition. She was awarded the 2003 Technical Prize from the Japanese Society of Kansei Engineering and she has given a TEDxStMartin talk (2012).

Pradipta Biswas is an assistant professor at the Centre for Product Design and Manufacturing of the Indian Institute of Science. His research focuses on user modeling and multimodal human-machine interaction for aviation and automotive environments and for assistive technology. He set up and leads the Interaction Design Lab at CPDM, IISc. Earlier, he was a senior research associate in the Engineering Department, a research fellow at Wolfson College, and a research associate at Trinity Hall of the University of Cambridge. He completed PhD in computer science at the Rainbow Group of the University of Cambridge Computer Laboratory and Trinity College in 2010, and was awarded a Gates-Cambridge Scholarship in 2006. He undertook a masters degree at the Indian Institute of Technology, Kharagpur. He conducted a course on HCI at the Indian Institute of Technology, Mandi, gave a guest lecture at the Indian Institute of Technology, Madras, and was a vice chairman of ITU-T Focus Group on Smart TV.

John Black, after a youthful career as a fine artist, doing painting and sculpture, has been developing software for more than 30 years. He has coded using a vast range of programming languages, from Z-80 assembly language to Prolog. He has founded and run several startup businesses developing business software and worked at large international corporations, such as Thomson Reuters, where he was both a software architect and a software development manager. Always striving to keep abreast of new technologies, he has run full-scale Bitcoin and Ethereum block chain nodes for several years, and worked with the Ethereum source code. As part of a commitment to furthering software standards efforts, he has worked with the Object Management Group (OMG, a worldwide standards organization) on a standard for employee time-recording data and worked with the W3C during the development of the resource description framework (RDF) standard for use on the Semantic Web. In 2006, he wrote the paper, "Creating a common ground for URI meaning using socially constructed Web sites."

Samantha Breslin is a PhD candidate in the Department of Anthropology at the Memorial University of Newfoundland (MUN). Trained initially as a computer scientist at the University of Waterloo, she then completed a master's degree in anthropology at MUN. Samantha's doctoral research is an ethnography of undergraduate computer science education in Singapore, exploring the "making" of computer scientists as (trans)national citizens and subjects in relation to computer science, world making, entrepreneurialism, and gender. Alongside this research, she has been working with Dr. Bimlesh Wadhwa at the National University of Singapore towards developing a gender and HCI curriculum to enable undergraduate computing students to explore how gender—and values more generally—are embedded in their programs, designs, and practices.

Noirin Curran received a PhD in applied psychology from University College Cork, and within psychology, her specialized area of interest is immersion in games. As part of her doctoral work, she used rigorous psychometric procedures to create the IMX Questionnaire, which measures the level of immersive response as experienced in a game session. Being fascinated by the experiences that can be prompted by modern media and technology, her current work in the game industry falls under the banner of HCI, where she investigates what game players do, and why, and promotes user-centered design based approaches. She has had the opportunity to carry out research involving a variety of advanced research methods, statistical work, and human-factors and usability-based methodologies in both academic and industry settings.

Clarisse Sieckenius de Souza is a full professor of the Department of Informatics of the Pontifical Catholic University of Rio de Janeiro (PUC-Rio). She received her doctorate in applied linguistics in 1987 from PUC-Rio's Department of Letters, with a thesis in natural language processing (NLP). In 1988, she joined the Department of Informatics, first involved in teaching and research on NLP and text generation and soon starting her lifetime work in semiotics and HCI. Clarisse is the creator of semiotic engineering, a semiotic theory of HCI for which, along with her colleagues and students, she developed specialized methods and models for interaction design. In recognition to her contribution to the field, she received the ACM SIGCHI CHI Academy Award in 2013 and the IFIP TC13 Pioneers of HCI Award in 2014. Her work has been published in over a hundred papers and she is the author or co-author of four books on semiotic engineering: The Semiotic Engineering of Human-Computer Interaction (2005); Semiotic Engineering Methods for Scientific Research in HCI (2009); A Journey through Cultures: Metaphors for Guiding the Design of Cross-Cultural Interactive Systems, and Software Developers as Users. Semiotic Investigations on (2013) Human-Centered Software Development (2016).

Jonathan Earthy works for Lloyd's Register where he coordinates the introduction of human factors and a human-centered approach into its products and the marine industry in general. His technical specialty is the assurance of the quality of human-centered design. He is an adjunct associate professor at the Australian Maritime Academy. He has participated in ergonomics and systems standards development since the mid-1980s and is convener of ISO TC159/SC4/WG6 human-centered design for interactive systems.

Julien Epps is an associate professor of signal processing at the School of Electrical Engineering and Telecommunications at the University of New South Wales (UNSW), Australia. He is also a contributed researcher with Data61, CSIRO, Australia. He holds a PhD (2001) in signal processing, and is the author or coauthor of over 200 journal articles, refereed conference papers and book chapters. He is currently is serving as an associate editor for *IEEE Transactions on Affective Computing* and for the human-media interaction section of *Frontiers in ICT* and *Frontiers in Psychology*, and is a member of the advisory board of the *ACM International Conference on Multimodal Interaction*. His primary research areas are speech and behavioral signal processing, cognitive workload, emotion and mental state recognition, and machine learning for human-computer interaction.

Jerry Alan Fails is an associate professor in the Computer Science Department at Boise State University in Boise, Idaho. His primary area of research is Human-Computer Interaction, with a focus on technologies that promote children's creativity, activity, mobility, collaboration, and exploration of the world around them. He has been actively designing technologies with and for children utilizing—and further developing—participatory design methods for children since 2003.

Peter Flynn manages the Academic and Collaborative Technologies Group in IT Services at University College Cork (UCC), Ireland. He trained at the London College of Printing and did his MA in computerized planning at Central London Polytechnic (now the University of Westminster). He worked in the United Kingdom for the Printing and Publishing Industry Training Board as a DP manager and for the United Information Services of Kansas as an IT consultant before joining UCC as project manager for academic and research computing. In 1990, he installed Ireland's first Web server and now concentrates on academic and research publishing support. He has been Secretary of the TeX Users Group, deputy director for Ireland of European Academic and Research Network (EARN), and a member both of the Internet Engineering Task Force (IETF) Working Group on HTML and of the W3C XML SIG; and he has published books on HTML, SGML/XML, and LaTeX. Peter also runs the markup and typesetting consultancy, Silmaril, and is editor of the XML FAQ as well as an irregular contributor to conferences and journals in electronic publishing, markup, and humanities computing, and a regular speaker and session chair at the XML Summer School in Oxford. He did his PhD in user interfaces to structured documents with the Human Factors Research Group in Applied Psychology in UCC.

J. J. Guajardo has over 15 years of experience as a User Researcher. After earning a bachelor's degree in psychology from Northwestern University, he received his PhD in developmental psychology from the University of Chicago in 2002. Immediately after, J. J. came to Microsoft to work on Xbox games. Following his stint in the gaming world, he worked with a number of products at Microsoft, including Encarta and Office for Mac. From 2007–2009, J. J. lived and worked in Copenhagen, Denmark, conducting user research for Microsoft Dynamics. In 2009, he returned to the United States and spent 2 years working in the Windows design and research group. In 2011, he returned to Xbox Research to work on kid-focused products and nongame entertainment efforts. He currently supports the Turn 10 franchise, working on the latest versions of the premier racing titles Forza Motorsport and Forza Horizon.

Mona Leigh Guha is the director of the University of Maryland's Center for Young Children. She has also been the interim director of the University of Maryland's Human-Computer Interaction Lab (HCIL), as well as managing director of KidsTeam, a team of adults and children who work together to design innovative technology for children. Her research has focused on working with young children as design partners and investigating the cognitive and social experiences of children who participate on a design team.

Daniel V. Gunn is a senior user research operations lead within Xbox Research. His team is responsible for driving the tech and facilities, people and process, and tools and infrastructure that empower world-class user research. Daniel's background is firmly seated in psychological research methods, statistics, and human behavior. He received his PhD in experimental psychology with an emphasis on human factors from the University of Cincinnati (UC) in 2002. Daniel has presented at several human-factors-related conferences and has published in the *American Journal of Psychology* as well as the *Journal of the Human Factors and Ergonomics Society.* In addition, he has coauthored articles and book chapters on the methodologies utilized within Xbox Research to improve games in development. He has worked on several Microsoft Studios titles across a variety of genres and platforms including installments in the Forza Motorsport series across Xbox and Xbox 360 as well as PC titles such as Rise of Nations: Rise of Legends and Viva Piñata PC. Daniel is always looking for ways to generate new types of insight for game designers leveraging innovative research methods.

Jerome R. Hagen is a senior user researcher in Xbox Research at Microsoft. He currently leads research on Minecraft and has led research on game franchises including Halo, Fable, Crackdown, Project Gotham Racing, and Phantom Dust. His background is in social / cognitive psychology and he also leads training for researchers on the Xbox team. He has led Team Xbox LGBTQ and is part of Xbox's focus on Gaming for Everyone to help make Xbox a place where everyone is welcome, respected, and supported.

Theodore D. Hellmann is a product manager at Splunk, where he works to support and expand its developer ecosystem. His work focuses on making sure third-party developers are provided with the tools and guidance to build Splunk Apps that ingest and store huge amounts of data, then make that data easy to use and understand. In his previous life in academia, he was a member of the Agile Surface Engineering Lab at the University of Calgary, where his research interests included test-driven development of graphical user interfaces, interactive and proxemic emergency operations planning, and interaction with / visualization of large-scale data.

Deborah J. O. Hendersen is a senior user researcher working on Xbox research at Microsoft. She received her doctorate from Stanford University in 2008 in cognitive psychology, where her work focused on understanding the differences between fiction and nonfiction. She is currently the user research lead for the Global Publishing Studio, and has supported numerous titles such as Double Fine's Happy Action Theater, Undead Lab's State of Decay, and Remedy's Quantum Break. Periodically, Dr. Hendersen shares out her work, most recently at the Games User Research Summit (https://www.youtube.com/embed/eWt3iEbTOX4) and Game Developers Conference (GDC).

Gabriela Jurca has experience as a research assistant in Frank Maurer's agile surface engineering lab, where she studied the integration of agile and user experience design. She has also completed her Masters of Computer Science at the University of Calgary, where she studied the application of data mining and network analysis to cancer research. Gabriela is currently working as a developer in the industry.

Todd A. Kelley received his doctorate from Johns Hopkins University 2007, where he studied how task practice can affect attention and distraction. He then worked as a postdoctoral fellow at University College London and University of California Davis, studying attention and perception using fMRI, EEG, and TMS. Todd joined Xbox Research in 2012, where he worked on latency perception, eye tracking, and biometrics. He led the refinement of the group's eye tracking methods, making it suitable for large scale studies with quick turnaround. He has also led the user research efforts on Dead Rising 3, Rise of the Tomb Raider (RotTR), and Dead Rising 4, and helped with Sunset Overdrive. Todd is especially proud of how RotTR made extensive use of the narrative testing techniques pioneered by this team.

Jurek Kirakowski comes from a practical computer science and psychology background. His speciality is quantitative measurement in human-computer interaction and he has contributed numerous books, articles, and workshops to this theme. His major research goal has been to show and indeed prove how the quality of use of information technology products can and should be quantitatively measured in an objective manner in order to support the management of developing good products. His original PhD was in speech perception, and he was one of the founder members of the Edinburgh School of Epistemics in the 1970s. He participated in artificial intelligence projects at the university at the time, in particular on state space representations and problem-solving strategies. Dr. Kirakowski took up the position of college lecturer in University College Cork in 1978. In 1984, he founded the Human Factors Research Group and was soon involved in one of the earliest of the Comission of the European Communities (CEC)-sponsored projects involving human factors: the Human Factors in Information Technology (HUFIT) project (ESPRIT 385). Dr. Kirakowski has since worked on numerous projects part-funded by the CEC on usability measurement and evaluation as it was known at the time, and user experience, which has latterly become popular. He has also worked as a technical reviewer for the CEC and as a consultant on technical panels in the areas of software metrics, measurement, and usability. He has wide experience as a user experience consultant to the IT industry. In 1999, he became statutory (senior) lecturer at University College Cork, and retired in 2015. Since his retirement he has developed the user experience solutions project featuring questionnaires and other resources he and his students have developed over the years. This project is housed at the uxp.ie website.

Pat Langdon is a principal research associate in the Cambridge University Engineering Design Centre (EDC) and lead researcher in inclusive design. His past research has examined the psychological reality of certain artificial intelligence-based theories of computer vision and neural-network algorithms for robot control as well as computational support for engineering design. He is currently working in the areas of modeling inclusive interaction, particularly vision, learning, movement, and cognition for inclusive design and computer assistance for motion impaired interface use. Pat is author and lead researcher responsible for a number of projects including:

- multimodal interfaces for adaptively creating inclusive interfaces for mobile device (IU-ATC) and interactive digital TV (EU GUIDE);
- human machine interfaces as applied to automotive displays and controls using signal processing for gestural and pointing intent (MATSA, MATSA2);
- inclusive human machine interfaces for the future car (CAPE iHMI project);
- haptic interfaces for touch audio devices (JLR TADHADIA);
- psychological models of latent variables in signal processing and automative machine learning (CAPE DIPBLAD).

He is currently coinvestigator and lead researcher for the successful bid for the joint EPSRC/Jaguar Land Rover-funded programme, Towards Autonomy—Smart and Connected Control (TASCC), Designing Autonomy in Vehicles (HI:DAVe) consortium. This is a joint collaboration between the EDC and the University of Southampton; running until 2019, which will conduct research into, and optimize practical solutions for, the difficult problem of how to interface drivers with automated vehicles. Dr. Langdon is a member of the ethics panel of the Cambridge School of Technology including the computer lab, the Engineering Department, the Judge Institute, and other labs including the Cavendish Lab. He has been instrumental in the development of the working practices, principles, and governance of this panel over several years. He has been external examiner for the Kings College London and Guy's Hospital intercollegiate MSc in assistive technology and teaches human factors on the MSc: safety engineering in the nuclear, rail, and aerospace industries, at Lancaster University.

I. Scott MacKenzie's research is in HCI with an emphasis on human performance measurement and modeling, experimental methods and evaluation, interaction devices and techniques, text entry, touch-based input, language modeling, accessible computing, gaming, and mobile computing. He has more than 160 peer-reviewed publications in the field of HCI (including more than 30 from the ACM's annual SIGCHI conference) and has given numerous invited talks. In 2015, he was elected into the ACM SIGCHI Academy. That same year he was the recipient of the Canadian Human-Computer Communication Society's (CHCCS) Achievement Award. Since 1999, he has been associate professor of computer science and engineering at York University, Canada. Home page: http://www.yorku.ca/mack/

Frank Maurer is the head of the Agile Software Engineering (ASE) group at the University of Calgary. His research interests are immersive analytics, multisurface systems, engineering analytics applications, and agile software methodologies. He served as the principal investigator of the NSERC SurfNet strategic network. The SurfNet Network was a Canadian research alliance of academic researchers, industry partners, and government collaborators. The goal of SurfNet was to improve the development, performance, and usability of software applications for surface computing environments: nontraditional digital display surfaces including multi-touch screens, tabletops, and wall-sized displays. He served as associate vice-president (research) and special advisor for entrepreneurship and innovation for the University of Calgary. He is cofounder, CTO, and board member at VizworX (www.vizworx.com).

Atsushi Nakazawa is an associate professor in the Department of Infomatics at Kyoto University. He received his doctorate from Osaka University in 2001 in systems engineering. Afterwards, he worked in the Institute of Industrial Science, University of Tokyo, and then in the Cybermedia Center, Osaka University. From 2013, he joined Kyoto University. His research interests are in human behavior/mental analysis using computer vision, eye tracking, eye imaging, and motion capture systems. Dr. Nakazawa received the best paper award in the International Conference on Virtual Systems and Multimedia (VSMM2004) and Japan Robotics Society (RSJ). In 2016, his paper was selected as a "spotlight on optics" from the Optics Society of America (OSA). His recent interests are corneal reflection and biosignal analysis for affective computing.

Tim A. Nichols has led user research on game and experience development across a wide range of platforms and user interfaces, including Xbox, Xbox Kinect, HoloLens, and VR. He currently leads research teams on Windows app development and on mixed reality platforms and experiences. He received his PhD in engineering psychology from Georgia Tech.

Christian Nitschke received a Diplom (MS) in media systems from the Bauhaus Universität Weimar, Germany in 2006 and a PhD in Engineering from Osaka University, Japan in 2011, where he continued as a postdoctoral researcher. In 2013 he joined the Graduate School of Informatics, Kyoto University, Japan as an assistant professor. Since 2016 he has been a system engineer at Bosch Sensortec GmbH, Germany, working on vision-based interaction techniques. His interests include computer vision, computer graphics, display and projection technologies, and HCI.

Kent L. Norman is an associate professor in the cognitive area in the Department of Psychology at the University of Maryland. He received his doctorate from the University of Iowa in 1973 in experimental psychology. He is the director of the Laboratory for Automation Psychology and Decision Processes (LAPDP) and is a founding member of the Human/Computer Interaction Laboratory (HCIL, http://www.cs.umd.edu/hcil) at the University of Maryland. His research on judgment and decision making and problem solving as they pertain to human/computer interaction and cognitive issues in interface design is reported in *The Psychology of Menu Selection: Designing Cognitive Control at the Human/Computer Interface* (1991). Dr. Norman is the developer of HyperCourseware™, a Web-based prototype for blended learning. He is coauthor of the QUIS: The Questionnaire for User Interaction Satisfaction, licensed by the university to academic, corporate, and government usability labs. He is the author of *Cyberpsychology: An Introduction to Human-Computer Interaction* (2017) and the author or coauthor of over 80 journal articles and book chapters.

Randy J. Pagulayan is one of the first pioneers of the games user research discipline. As the director of Xbox Research, Randy leads a team at the forefront of interactive entertainment experiences at Microsoft across games and the Xbox platform. Previously, he has led research efforts on numerous blockbuster video games and franchises, including Age of Empires and Halo. He has also coauthored book chapters on user-centered design in games and testing methodologies, has given numerous talks and keynotes internationally, and has been featured in Wired and National Public Radio. Prior to joining Microsoft, Randy has published in several scientific journals, including Journal of Experimental Psychology, Brain Research Bulletin, and Human Movement Science. Randy has a BA in psychology from the University of Maryland, and a PhD in experimental psychology from the University of Cincinnati.

Bruce C. Phillips joined Xbox Research in 2001. He is particularly interested in counting things, visualizing things, and using data collected through telemetry systems to advance design intelligence as a tool to improve players' experiences. In 2003, he helped develop the first data-enriched telemetry system, which later became known as TRUE instrumentation, for use during the development of the game Voodoo Vince. Since then, the TRUE system has been used on dozens of game titles across a variety of genres. Prior to Microsoft, Bruce received a BA in psychology from Carleton University and a PhD from the University of Victoria.

Dave Randall is senior professor in the Department of Informatics at the University of Siegen, Germany, and visiting professor at the Linnaeus University, Sweden. He has published seven books and a large number of peer-reviewed papers in the area of computer supported cooperative work (CSCW) and HCI. His substantive interests lie in the use of qualitative methods, and notably ethnographic approaches, for deriving design insights in a variety of different contexts. His most recent book, *Choice* (with Richard Harper and Wes Sharrock, 2016), is an examination of disciplinary assumptions in relation to the problem of decision-making.

Sirpa Riihiaho works as a postdoctoral researcher at the University of Helsinki in the Discovery Research Group in the Department of Computer Science. This group works on artificial intelligence and data science, especially on computational creativity and data mining. Previously, she worked as a senior lecturer in Aalto University in the Strategic Usability Research Group (STRATUS) in the Department of Computer Science. This group does research and provides teaching in usability engineering and user-centered design. Her research and teaching has focused on the methods for user-centered product development, especially on usability evaluation and usability testing methods. Her doctoral thesis, *Experiences with usability testing: Effects of thinking aloud and moderator presence* (2015), combined an extensive literature review with 22 years experience of usability testing, covering 143 usability studies.

Mark Rouncefield is a reader in the Department of Computing at Lancaster University, United Kingdom. He is well known as an ethnographer in organizational and community contexts and has published extensively on these and other themes. He is author, along with Dave Randall and Richard Harper, of *Fieldwork for Design,* and the editor of *Ethnomethodology at Work,* and *Ethnomethodology at Play* with his colleague, Peter Tolmie.

Takaaki Shiratori is currently a research scientist at Oculus Research. He received a BE, an ME, and a PhD in information science and technology from the University of Tokyo in 2002, 2004, and 2007, respectively. He was previously with the visual computing group at Microsoft Research. Prior to that, he held postdoctoral researcher positions at Carnegie Mellon University and Disney Research Pittsburgh. His research interests lie in computer graphics, computer vision, and human-computer interaction, with a particular interest in user interfaces for character animation and character interaction. He won the Best Paper Award at IEEE 3DUI in 2013. He is currently an associate editor of *Computer Animation and Virtual Worlds.* He was the program co-chair of the SIGGRAPH Asia 2015 Technical Briefs and Posters Programs, and has served as a program committee member for several computer graphics conferences, including ACM/EG SCA, ACM I3D and Pacific Graphics.

Jakob Grue Simonsen is professor of computer science at the Department of Computer Science, University of Copenhagen (DIKU). His primary research areas are computability and complexity theory, human-computer interaction, and information retrieval.

Mark Springett is a member of the Interaction Design Centre and the Design-for-All research group at Middlesex University. He was Vice-Chair of COST Action IC0904 "Towards the Integration of Trans-sectorial IT Design and Evaluation" between 2009 and 2013. He has over 30 years' experience of working in HCI in academia and industry, with an emphasis on design for specific user populations including those with disabilities. He has a specialist interest in the evaluation and modeling of user experience, and factors affecting acceptance and takeup of new technology. He is currently leading the EU Erasmus-funded project "Gameplay for Inspiring Digital Adoption," which is concerned with the potential of novel interactive technologies to improve quality of experience and digital engagement of older citizens.

Ana Tajadura-Jiménez is a Ramón y Cajal research fellow at the Department of Informatics of Universidad Carlos III de Madrid (UC3M) and an honorary research associate at the Interaction Centre of the University College London (UCL). She received her PhD in applied acoustics from Chalmers University of Technology, Gothenburg, Sweden, in 2008. Her research is empirical and multidisciplinary, combining perspectives of psychoacoustics, neuroscience, and HCI. In her PhD studies she adopted an embodied perspective to investigate auditory-induced emotion and multisensory integration processes. In her current research she focuses on the use of body-sensory feedback to change the mental representation of one's own body, and her studies pioneer the use of sound for producing these changes. She coordinates the research line Multisensory Stimulation to Alter the Perception of Body and Space, Emotion and Motor Behavior, and is principal investigator of the MagicShoes project (www. magicshoes.es), which is developing wearable technology that integrates body sensing and sensory feedback. Dr. Tajadura-Jiménez has published more than 50 papers and book chapters. Her work has been featured in the media worldwide. A highlight is the article on her project The Hearing Body, which appeared in 2015 in *New Scientist.*

Harold Thimbleby has been researching HCI since the 1970s, particularly seeking rigorous principles of design that will help developers design safer and more effective user interfaces. He has particularly focused on healthcare HCI where the problem is neither UX nor ease of use but that poor design harms and kills people unnecessarily. Healthcare computer systems have multiple types of users: patients, clinicians, hospital managers, IT managers, paramedics, even the police, and all with very complex tasks and very different requirements for IT support. Healthcare developers have an overwhelming task. To design well, you have to take a more abstract view of what interaction is. He is professor of computer science at Swansea University, Wales, and is an honorary fellow of the Royal College of Physicians Edinburgh and of the Royal College of Physicians London, where he is expert advisor for IT. He has published over 350 refereed papers and is a well-known speaker. His website is http://www.harold.thimbleby.net

Aleksander Väljamäe is an associate professor in physiological computing in the School of Digital Technologies at Tallinn University, Estonia. He received his PhD in applied acoustics at Chalmers University of Technology, Gothenburg, Sweden, in 2007, focusing on multisensory perception in motion simulators, especially, on auditory induced illusory self-motion. His current psychophysiology research concerns how audiovisual media influence humans on the perceptual, cognitive, and emotional levels, with particular stress on the novel methods for diagnosis and treatment of various brain disorders (e.g. depression, migraine) and new applications (brain-computer interfaces, neurocinema, neurotheater). Dr. Väljamäe also participates actively in art and science projects, for example his technical directing of the "Multimodal Brain Orchestra" performance in 2009, Prague, or directing neurotheater performance "Demultiplexia" in 2017, Brussels. He is the author or coauthor of over 30 journal articles and book chapters.

Bimlesh Wadhwa is a senior lecturer in the School of Computing at the National University of Singapore (NUS) where she has been a faculty member since 2000. She completed her BSc (physics, 1983), MSc (physics, 1985), and PhD (computer science, 1990) at Delhi University, and has an MTech in software engineering. She has also published widely on these topics. She has served on many conference and workshop organization and program committees, as well as on many hackathon and coding competition committees.

Xiaoge Xu is a full professor and the founding director of Center for Mobile Studies at Xiamen University, Malaysia. Since he founded Mobile Studies International in 2012, he has been passionately promoting and conducting mobile studies. He is the editor of two mobile-related books: *Interdisciplinary Mobile Media and Communications: Social, Political and Economic Implications,* and *Handbook of Research on Human Social Interaction in the Age of Mobile Devices.* As an advocate and scholar of mobile studies, he has initiated quite a few projects including the Mobile Studies Congress, the Mobile Museum, Mobile Summer Schools, Mobile Workshops, Mobile News Experience, Mobile Healthcare Experience, Mobile Learning Experience, and Mobile Competitions.

Acknowledgments

We, Kent and Jurek, would like to dedicate this handbook to our loving and devoted wives, Karen and Maìre Doṁnat, respectively, who have long held our sanity and wiped our brows when necessary; to one another who have helped each other to overcome the pitfalls of life and to be academic and spiritual brothers; to our many colleagues in computer science, engineering, psychology, and other disciplines who have worked hard and long to bring us to this point in history; and above all to the Creator of all things visible and invisible, which includes the human-computer interface and beyond.

Such a large and ambitious project would have been impossible without the continued help and dedicated assistance of the wonderful team at Wiley-Blackwell from the editor of the psychology series, Andrew Peart, who first anticipated the need for this handbook to the many who helped to shepherd it through to completion: our project editors in order, Karen Shield, Roshna Mohan, and finally Silvy Achankunji; our copyeditor, David Michael, whose sensitive touch and appreciation of the finer points of APA style were refreshing and educational; our production editor, Kumudhavalli Narasimhan, who brought this project to its completion; and finally, for our cover design, Monica Rogers.

Introduction: Human-Computer Interaction Yesterday, Today, and Tomorrow

Kent L. Norman and Jurek Kirakowski

A Very Human Fascination

A characteristic of humans is that we have an enduring fascination with tools. Humans construct tools to be useful and to serve a particular purpose. But tools are also objects with other properties. They are works of art, they are possessions, and they customized. Even from the earliest tools of our ancestors, such as the Paleolithic flint scrapers, we see that humans not only fashioned tools to be usable but also fashioned them in a way that the maker could take pride in them. These artifacts are given a value beyond being merely functional (Berleant, 2007). Moreover, tools are extensions of the human body and the mind. As such, they spawn metaphors of the structure and function of our interactions with tools. In fact, many expressions such as "impression," "smoothing over," and "clean slate" may be considered as cultural back references to an impressive piece of classical ancient-world technology that prompted Plato to use as his model of the human memory the wax tablet (see Plato's *Theaetetus,* 191c *et seq*., http://www.perseus.tufts.edu/hopper/text?doc=Perseus%3Atext%3A1999.01.0172%3Atext%3DTheaet.%3Asection%3D191c). No doubt the reader will imagine many more collusions, if not collisions, between technology and the human mind. The song by Bob Geldof, "I Don't Like Mondays," with its use of the metaphor of the overloaded silicon chip, continues the tradition of Plato (Clarke, 1979).

The origins of stored-program computing devices are obscure, with many inventors proposed for the laurels (e.g., George Boole, 1815–1864; Herman Hollerith, 1860–1929; Claude Shannon 1916–2001). But it was not until the 1960s that it became clear that organizations, whether government, business, or industry, could benefit from the appropriation of information technology. Then, in the 1970s, the first serious efforts began to be made to tailor computer technology to the average human user as there is always a marked shortage of engineers in the world. Lavington (1980) gives a fascinating account of how computing entered the British corporate market in those years. If a computer could do practically anything, then it was an obvious step to create computer programs that translated the needs of the human into the instructions necessary for making the machine satisfy those needs. It is worthwhile remembering that the acronym of the programming language FORTRAN stood originally for "FORmula TRANslator,"

The Wiley Handbook of Human Computer Interaction, Volume 1, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

and FORTRAN was heavily tipped as the ideal way for a mathematician to enter mathematical formulas into a computer.

Originally, our needs were modest and closely tied to the things that computers were originally designed to do, such as scientific and financial computing, record keeping, and even text processing. But when the possibility of geographically linking distant computers through telecommunications networks took hold, and the earliest information transfer protocols were proposed, the concept of the computer as a device totally changed. The idea of connecting humans to humans was born and envisioned by Licklider and Taylor (1968), and enabled through the ARPANET in the United States, and at the same time explored in the 1969 undergraduate computer science classes at Edinburgh University, which were already learning how communications protocols worked. Our enthusiasm for these technologies was encouraged by Professor Sidney Michaelson, who was at the time working on the groundbreaking concept of EMAS: the Edinburgh Multi-Access System, which allowed many users to use a geographically distributed set of computing resources. Communication by electronic mail (or e-mail as it was quaintly called in those days), online chatting, the sharing of documents by file transfer, and the invention of the multiuser dungeon (Wisner, 1990) soon followed. However, it was not until computers were affordable by the average citizen that the real evolution of computers as communications devices began.

The Expanding and Encompassing Interface

We can identify roughly three generations of popular human-computer interfaces. The first-generation devices and interfaces were modeled on their mainframe teletype ancestors, and consisted of green screens and keyboards. More technically advanced owners were able to network their computers to bulletin boards and chat forums, and many young people got their first taste of interacting with computers through BASIC, which on many so-called microcomputers, acted as both a primitive operating system and an even more primitive coding language. BASIC is still with us and, as one of us, namely the second editor of this handbook, predicted long ago, BASIC was too good an idea to die. It has continued to develop, and still does to this day, having first swallowed structured programming and then later on the object-oriented technologies of writing programs (Kirakowski, 1988).

The second generation of human-computer interaction (HCI) saw much more sophisticated screen technology, the advent of pointing devices such as a the mouse and the trackball, and the introduction of graphical user interfaces (GUIs) with windows, icons, menus, and pointing devices (WIMP). It was at this point that the first editor of this handbook was inspired to turn from the study of the cognitive processes of judgment and decision making to the study of menu selection at the human-computer interface (Norman, 1991).

At this time, vast increases in the amount of storage space available made it possible to store digitized versions of music, photographs, and films. Although the operating systems of such computers were not explicitly geared to communications networking (or to avoiding the perils attendant on such technology being widely available), it became possible to connect computers to networks such as the World Wide Web.

At this point, computer users started to enjoy the social advantages of connections between far-flung regions of the world thanks to the hypertext transfer protocol concept pioneered and promoted by Tim Berners-Lee and his colleagues in the late 1980s and early 1990s.

We are now arguably in the third generation of HCI. Computing devices are expected to be connected wirelessly and computing power is distributed in the cloud. We expect to be able to use the communications channels open to us for virtually no payment, at any time of the day or night, and we expect to be able to keep them in our purses or the back pockets of our jeans. Some of us expect to have access to enormous immersive screens and sound systems we feel we could walk into, and gestural devices that actually make us part of the action. Virtual reality, augmented reality, and location services are fundamental to this generation.

We use these wonderful devices for social purposes—for reaching out to, and making contact with, other humans. The end goal of such activities as buying an airplane ticket or smart new leather jacket is still ultimately social. Underpinning all of this are immensely deep layers of technology that no one single human could any longer possibly understand. Similarly, the costs involved are quite astronomically mind boggling. With regard to technological complexity, we've become accustomed to this, although many of us just don't believe in or trust the thin tissue on which we tread.

But, as the conclusion to Landauer (1995) suggests, cost is not an issue. Organizations will make money from the users involved, although what precisely such businesses are selling is not always clear to the average user, and only gradually is the industry becoming aware of the severe consequences of issues such as security and personal identity. Products that once sold at a fixed price, can become variable in price depending on who the purchaser is and their purchasing record.

Human-computer interaction has had a long history in a short span of time. Much has happened since the introduction of the MITS Altair 8800 in 1974. The interface and interaction between the human and the computer have changed with increasing velocity and spectrum of trajectories. The interface is everywhere (ubiquitous and mobile); the interface is visual (watching and showing); the interface is conversational (talking and listening); and the interface is smart. In retrospect it has taken us less than half a century to begin to fashion flint (well silicon) that can be used to provide a rich personal and social experience as well as being useful.

The human-computer interface was first conceived as that point at which input was received from the user and at which the computer output information to the user—namely, the computer screen and the keyboard and mouse. This interface still exists, and is not likely to go away, but innovation in HCI has opened many new channels, surfaces, and modalities, and it continues to expand.

In the past, the interaction between computers and humans was limited in time, quantity, and quality. Today, we are almost always interacting with computers and networks; the sheer quantity of things that humans and computers do together is huge; and the quality of the interaction is coming close to matching human perceptional and cognitive abilities. The interface is beginning to encompass all human activity. "Cyberpsychology" has become the term for this overlap of human activity and computer processing (Norman, 2017).

Global Reach

The field of HCI has likewise expanded globally from researchers primarily in the United States and the European Union to around the world. Some commentators have seen it as an outgrowth of the action-orientated discipline of ergonomics, popular in the 1940s, so giving it the somewhat oxymoronic label *cognitive ergonomics* (Budnick & Michael, 2001). However, to our ears such a label emphasizes only the tool, and not the use to which the tool is put. The authors contributing to this handbook see the tools, of course, but their emphasis is on the way the tools are put to human ends. The touch of HCI has expanded from the technically orientated pioneers to all sectors of humanity: rich and poor, gifted and disadvantaged, those who relish technical sophistication to those who prefer to see technology as transparent. The power of HCI has been applied to the human activities of working together, teaching, and enjoying ourselves in company, enabling things we could never have imagined in the early days of computing.

Organization

Our vision for this handbook was that the future of information technology is social. Although there is a need continuously to examine and improve the basic interface surfaces (screens, touch pads, motion sensors, audio channels, voice recognition, and indeed more direct neural interfaces), the real advances of the future, which we have barely begun to see, are in the revolution that this technology will make to our ways of interacting with each other.

We have therefore to envision new ways of building the products that will correspond to the vastly expanded possibilities the technology offers us. Very often, in our professional activities, when talking to the intended users of the technology, we have to remind them, don't be limited by what *you* think the technology can or can't do. Tell us what you *want* the technology to do for you. We might not get there right away, but you'll see us heading in that direction. When talking to designers, on the other hand, both of us have frequently found ourselves saying, hey, that technology is fine and dandy, but what on earth can it be used for? Will anyone use it? Or are you just proposing to put it out there because it's fascinating for you, the technologist?

Well, the reverse of the stories is also true: often, ordinary people can't see the possibilities being opened out, and so their vision is limited to what they know. Often, the sheer excitement of the new technology will fire the imagination of countless end users who will adopt it enthusiastically, no matter how difficult it is to use (older readers will remember the craze for texting with the characters of the alphabet mapped onto the numeric keys from zero to nine on a pocket handheld device).

So how do we do this? How do we manage the design of our lovely gadgets?

In Volume 1, we go in a top-down manner considering the temporal order of creating interfaces from overarching design issues (Part I) to the actual process of design (Part II) and from factors of evaluation (Part III) to methods of evaluation (Part IV). Volume I ends with a consideration of the end user from input to output (Part V).

And what will be the effect of this technology on us humans?

Volume 2 opens with the interface (Part VI) and the interactions that take place there (Part VII). The remainder of Volume II is more-or-less bottom up, dealing with accessibility and special needs of some users (Part VIII), the social aspects of users (Part IX) and communities (Part X), and finally with the design and implementation of a number of specific applications (Part XI).

The Future

We hoped not to create a retrospective body of work from the contributions of our outstanding collaborators—those men and women working on the raw edges of making the technology work for people. Of course, as has often been pointed out, those who ignore the past are doomed to repeat the same mistakes in the future. A handbook should not be a collection of recipes any more than it should be a dust-attracting volume of history. We have tried to make it true a *vade mecum,* a volume to have with us on our journey, which can be opened at any moment to give inspiring stories for all of us involved in human computer interaction for many years to come. So it's not just background material and current practices in HCI; our chapters also contain information about innovations likely to change the future of HCI and suggest, we hope, new ways of thinking.

These are exciting times. We hope that we have been able to stimulate a way of thinking about the future that will enable us to use the successors to those old flint tools and wax tablets to create truly wonderful clothes and artifacts, which will make possible and fashion the amazing personal interactions of the glamorous society of the future.

References

Berlant, R. (2007). Paleolithic flints: Is an aesthetics of stone tools possible? *Contemporary Aesthetics, 5.* Retrieved from www.contempaesthetics.org/newvolume/pages/article.php?articleID=488

Budnick, P., & Michael, R. (2001, June 11). What is cognitive ergonomics? [Rev. ed.]. Retrieved from https://ergoweb.com/what-is-cognitive-ergonomics/

Clarke, S. (1979, October). The fastest lip on vinyl. *Smash Hits,* 6–7.

Kirakowski, J. (1988). *Human-computer interaction: From voltage to knowledge.* Lund, Sweden: Chartwell-Bratt.

Landauer, T. (1995). *The trouble with computers, usefulness, usability, and productivity*, Cambridge, MA: MIT Press.

Lavington, S. (1980). *Early British computers.* Manchester, UK: Manchester University Press.

Licklider, J., & Taylor, R. W. (1968). The computer as a communication device. *Science and Technology, 76*, 21–31.

Norman, K. L. (1991). *The psychology of menu selection: Designing cognitive control at the human/computer interface.* Norwood, NJ: Ablex Publishing Corporation.

Norman, K. L. (2017). *Cyberpsychology: An introduction to the psychology of human-computer interaction* (2nd edn.). Cambridge: Cambridge University Press.

Wisner, B. (1990). *A brief history of MUDs.* Retrieved from groups.google.com/forum/#!msg/alt.mud/m2jy5tBY9Lo/um6mxNXFwaAJ

Part I
Design Issues

1

Interactive Critical Systems and How to Build Them

Harold Thimbleby

Introduction

We all know many frustrating examples of interaction, which feel even worse because often they need not have been designed that way. Many frustrations are either simply tolerated or passed by because there is nothing too important for the user riding on the result. Some frustrations are noticed when the user is trying to do something important; then, often, the urgency of the ongoing task interferes with trying to understand the frustration—and certainly interferes with any motivation to work out some helpful things to feed back to the designers. But when the tasks are life critical—like flying planes or giving patients radiotherapy—users have even less time or mental capacity to think about the causes of their frustrations with the interaction. Ironically, for some systems, user frustration may be also experienced with logging in and similar communication problems, yet until one is successfully "registered," complaining and raising bug issues is impossible.

If we did tell the designers, their eyes and ours would glaze over well before anything happened. By the time users call for help, they have probably had a long and frustrating experience, and there is probably a large conceptual gulf between them and their experience and the programmers who can fix the problems and the issues they understand. So instead of complaining after frustrations, we need principles—much higher level ways of talking about problems—so the frustrations can be avoided.

We do not often tell a woodworker to remove this splinter and that splinter—and the one over here, or this one… and I haven't time to tell you about *these ones* because I need to use the table right now! Instead, we expect woodworkers to use a process that *avoids* splinters. Sanding and polishing perhaps. Splinters, like bugs, are a symptom that a process that has gone wrong; identifying and fixing splinters one by one is not the best solution. We do not expect to have to tell professional woodworkers about splinters, and we certainly don't expect to have to wait to identify each one until something serious has gone wrong—it would be sufficient to say the wood just needs smoothing off.

The Wiley Handbook of Human Computer Interaction, Volume 1, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

The Air Inter Flight ITF148 Lyon to Strasbourg crash on 20 January 1992 had multiple causes. The aviation safety accident report cites a crew workload peak, and the fact that the crew did not notice an excessively high rate of descent until too late. The excessive rate of descent may have been partly caused by a failure in the flight control unit (FCU) to enter correctly into vertical speed (VS) mode instead of flight-path angle (FPA) mode, or by the pilots being confused as to which mode they were in. The pilots entered 33 intending a 3.3 degree descent angle, which the autopilot treated as a descent rate of 3,300 feet per minute. Both would have been displayed as "–33." As a result of this accident, Airbus made some design improvements to the FCU giving the digital VS mode readout four digits and the FPA readout just two. Furthermore, 34 safety recommendations were issued by the French BEA (Aviation Safety Network, 1992).

Some calculators recently on the market implement an unexpected decimal-point design feature. When users key in numbers, the keyboard provides a key click (audible feedback) to confirm the keys have, in fact, been pressed. If the decimal point is pressed more than once, there is a still a key click for each press (providing confirmation that the decimal point has been pressed) yet nothing happens at all: it would seem that, once a number has a decimal point, further presses of the decimal points do nothing.

On other calculators, the decimal point behaves differently. On many, pressing the decimal point moves the decimal to the right—so if you press "2.5.6," the number actually entered would be "25.6." Two clicks of a decimal point while entering a number is certainly a user error. The behavior of these calculators fails to detect this user error, but they handle the error differently. Users, *if they notice,* will be surprised by the number actually entered. This may not sound very significant, and is most of the time an instance of the first kind of frustration outlined at the start of this chapter—something, if noted, to be passed by.

The design of calculators and the problems of their user interfaces is not usually a critical problem, although one may argue that they may be critical "in scale" so that a trivial frustration experienced by millions of users in an everyday situation should be as noteworthy as a big frustration experienced by very few in a safety-critical situation. However, there is the possibility that a mass-produced piece of equipment could itself be used in a safety-critical situation where the unexpected behavior of the equipment in reaction to a user error may not be noticed, and unlike in the everyday situation, serious consequences may ensue (Miller, 2013, raises the issue of the danger of relying on "COTS" or "commercial off-the-shelf" hardware and software in safety-critical situations for which it was not explicitly designed).

Such critical problems are not going to be solved by user-centered design; designers and users are unaware of them, or users are too busy trying to do their work, without spending time taking issue with poor design. Typical user-centered design evaluations are far too small and too short for problems like these to be registered as "statistically significant." Of course, user-centered design is important (Landauer, 1995) but it is clearly not the whole story. It can find some splinters but it cannot avoid them systematically—it is like wiping your hand over wood to see if it catches on any splinters within sight, rather than just preparing the wood properly in the first place.

Technical Debt

The concept of "technical debt" can be illustrated by the following anecdote.

I was trying to create a user account to send an email to a company because I wanted to raise a problem I had experienced a moment earlier, which had stopped my purchasing process.

You know how it goes: registering asks you for everything. Your gender must be supplied, as if it mattered to a computer. They want your date of birth—taken from a drop down menu of over 100 choices. I tried typing my date of birth (it happens to be 1955) and the menu selected 1959 (probably as the largest number starting 195). So they thought about keyboard shortcuts but did not implement them correctly.

I had to enter a password: well, I have my methods (a letter, a number I remember, and some punctuation to keep the computer happy if it needs some). But as soon as you start to type the password, it tells you something complicated—I think it said "password must be at least eight characters and enter at least two of the following: digit, lower case, upper case, punctuation." Hey! I'm trying to remember a password, not play a word game! Of course, my browser hasn't been told by the company's system that this field is a password, so the browser isn't going to help me by remembering it when I need it later.

Next I enter my phone number. The box says "Phone number (United States +1)," so I enter "+447525191956," and it says "it doesn't look right"! So I realize that "(United States +1)" is a "secret" drop-down menu, and I find United Kingdom (+44) in the menu—but now they have deleted the phone number I had just entered, so I have to enter it again (this time without the +44). I must say that if telephones can cope with international dialing, I don't really understand why a website can't get it right.

Finally, after filling in everything on this long form I hit NEXT and then it says "This site is temporarily unavailable due to maintenance. Please try again later."

And the "go back" button goes to a page that says the same thing. My form has gone! My details have gone! I put a lot of effort into my new password! The computer has forgotten everything, and tomorrow I will have to go through the whole process again. I thought I'd tell them some of my experience but you need an account to contact them. It is as if they have carefully thought this through. Maybe they don't want to hear from people who can't even log in?

There are many obvious design failings in this sadly familiar experience.

User-centered design probably won't help because the probability of experiencing the problems is too low to be significant (particularly if the company has performed small evaluations and got rid of what they see as the big problems and if their clientele is largely based in their own country). Their feedback system (if they use iterative design) presumes you can even get past the first stage, so it is already success biased. They will only receive feedback from people who can create an account and log in.

The company may lose a little money (i.e., profit from goods and services I will not be able to buy) but they didn't notice the costs to me. If they are rational, then, what they think they lose is less than the cost of doing the job properly: this is what is called *technical debt* (Allman, 2012). To save money at the design stage they are essentially getting users to pay extra later. Savings in design are paid back by users paying the interest on the design debt.

Solving Problems

The first step in solving a problem is to realize there *is* a problem, to identify the problem, and maybe give it a name so that people can talk about it. Alternatively, one can *dissolve* the problem, or even hope that it just goes away, which is evidently the preferred method—not many people are complaining, so it might be considered something to ignore rather than to solve.

Then, having identified the problem, we have to express it in a way that allows it to be solved, and ideally we need to find a way that leads to *consensus*. Companies or other organizations solving their problems and us thinking the problems are solved are different points of view! We need a method that we can (in principle) agree on.

Turning a problem into mathematics is a good way to do this.

Instead of arguing over how much money I owe you, we could write it down, add up the columns—and what columns of numbers do is and has been agreed for centuries. Once it is expressed mathematically, any problem is over half way to achieving consensus on its solution.

A well-known example (from Kahneman, 2012) is as follows. If a baseball bat and a ball cost a total of \$1.10, and the bat costs \$1 more than the ball, then how much does the ball cost?

The answer is \$0.10, right?

Wrong!! (It is fun to do this at a party.)

Instead, let's use mathematics. Call the cost of the bat x and the cost of the ball y. We are told $x+y=1.1$, and we are told $x-y=1$, and we want to know the cost of the ball. Mathematically, then, we want to know y, but unfortunately both equations with y in them have also got an unknown x in them, so we need to eliminate x from one of the equations we know; then we can arrange the equation to tell us what y is.

For example: we know $x-y=1$, so $x=y+1$. We can use that to substitute for x in the equation $x+y=1.1$, to get $y+1+y=1.1$. We simplify this to $2y=1.1-1$, which is $2y=0.1$, which is $y=0.1/2$, which gives us $y=0.05$. So, the ball costs \$0.05. Almost certainly, if you obtained the right answer to Kahneman's problem, you did something very close to this mathematics in your head: you used mathematics to think clearly. One of the interesting lessons here is that converting the problem to mathematics gives us several ways to solve the problem, and hence—if you wanted—several ways to check your answer, and because different people can solve problems in different ways, it provides an important way of achieving consensus and agreement.

The big picture of this chapter is that computer system engineering and design problems are *much* harder than working out the cost of a 5 cent ball, but the same principles of applied reasoning to get the right answer, and to know it is right—namely, mathematics—help us dramatically.

> Designing critical systems, particularly those involving human-computer interaction, is a lot more complicated and prone to error than Kahneman's simple problem.

Civilization has progressed to where it is today because we found some very powerful techniques for identifying problems and getting consensus on their solutions—primarily the scientific method combined with mathematics. There is a bit more to this standard formula: once we have solved problems, society needs to remember so we don't have to keep repeating the effort of solving problems (or even repeating the effort of noticing that there are problems to be solved). Technology is the ratchet that makes progress and keeps civilization from falling back. And good interactive computer systems do exactly that: they make the world a better place, and better than solving one problem at a time, they make solutions available to lots of people in lots of places. That is why it is so important to have good solutions based on good thinking.

In HCI we often forget about the need for reliable *reasoning* about design—that is, doing mathematics. Simply observing the world reliably (typically by doing empirical experiments followed by statistics to demonstrate that our answers are not due to chance variation) is not enough to develop reliable design principles for critical systems, let alone to develop principles that are understood and can be used by developers. Conventional HCI may find out that one user interface is significantly different from another, but it is a rare piece of research that goes on to try to find out the principles underlying *why*, so that the next designer can benefit from the insight on their project.

One reason for this limitation of empirical HCI is that good principles have to be general, and apply to new product designs. This means they must generally be mathematical principles. Fitts' law is one such example. However, HCI researchers, designers especially, psychologists, and programmers rarely come across and feel comfortable with mathematics. In fact, it is possible—as the real examples above prove—it is easy to do a "good enough" job ignoring these issues: so why invest time in learning and using them? Again, this is technical debt.

Thus, user interfaces seem to be an exception to good engineering; think of other areas, such as mobile phone communication cells, which we take for granted when we use mobile phones. It is incredibly complicated to get these things to work— but telecoms service providers employ competent engineers to design and build the networks. Not everybody needs to be competent, but in computing, sometimes incompetence can be passed off as technical debt.

What is the Key Problem with HCI?

Put simplistically, the user's model is not compatible with the engineer's model, so both user and computer "do things" but they do not necessarily do the "right things" from each other's point of view. Then a sequence of misunderstandings ensues, which will end in frustration for the user.

Put more precisely, the user's model has properties, and the engineer's model has properties, and the appropriate properties must be and remain compatible over time as the human and computer interact.

It is in fact rather hard to define what is meant by "appropriate properties" and "compatible." Humans and computers can learn, and models can change. Right now my laptop has the property that it is running off mains power; if the mains failed, this would change nothing, because the battery would seamlessly take over. So far as I can see, no property has changed. But if things stay this way for a few hours, my laptop will die and I will be able to do nothing with it.

What was a hidden property can become critical.

It works the other way too: the computer thinks I am "present" but I may let the cat out and, when I resume, the computer has "timed out" because it thinks it is possible I may be somebody else. I clearly am not (I never have been!) but the computer is guessing a property about who the user is. In this case it is wrong.

Model mismatch leads to well-known problems. Fortunately, most of the time we can get away with ignoring, working around, or dismissing these problems; they are usually irritations rather than disasters.

There are many mathematical ways of describing a relation of models; we will start with *abstraction*. When we use an abstraction, we talk about an object's properties and relations, ignoring exactly how the object works like that. If my laptop is working, I can abstract away from the details of its power supply. Unfortunately, the abstraction will break if the *implementation* of it leaks into the abstraction. My abstraction might better have been "my laptop works and its power is OK." Abstractions, if not well thought out, can become overcomplex and cease to be useful. The term *encapsulation* means the implementation of an abstraction will not leak and make the correct abstraction overcomplex.

Conversely, we talk of *implementation bias* when an abstraction assumes, includes, or exposes unnecessary additional information about how it is implemented. In the literature implementation bias is discussed in a technical way[1] but in this chapter we are concerned with the impact on users. Here are some examples:

- Housing and other Agencies often ask for documents written using a proprietary commercial format. In principle, they just want documents, but they ask for a particular implementation of documents, namely that provided by a proprietary format they have purchased. It is understandable that the agencies bought the implementation with all the unique features that make it competitive (and that will no doubt be upgraded to stay competitive in the market) but the end result is that you can only send them "documents" that work in the format of the current implementation.
- Abstractly, calculators perform arithmetic. But if I discover that "$2 + 3 \times 4 = 20$" and if I assume this unusual implementation rule works universally then, when using another calculator that does arithmetic properly, I will make mistakes because of my bias to the irregular implementation of arithmetic on the "unusual" calculator.
- Consider an organization that amends its database to track staff gender preferences, probably to help it comply with equality laws. The implementation provides three options: *male*, *female*, and *prefer not to say*. The implementation has to choose something from these three for everyone to start with, and so it forced my record to "*prefer not to say*," when the truth was I never knew I had been asked; I had never said I preferred not to say—I am happy to say I am a male (if asked). The implementation has failed to provide the option *user has not reviewed this*. Such a poor implementation of a user interface makes the data collected biased towards "*gender neutrality*." Moreover, some users might prefer to say something that is none of the implemented options—there are sensible cases that are neither male, female, nor prefer not to say. What conclusions can be made on the basis of such flawed data?

[1] For example, an abstraction might require a set but it may be implemented as a list. If somebody knows this, they can use their knowledge of the implementation to find out, say, the most recent element put in the collection. If somebody then improves the implementation, say replacing the list with a tree, the abstract set is still valid, but anything that relied on finding the last element will fail because of its now incorrect implementation bias.

- In the old days, we could fill in paper forms how we liked and as slowly as we liked. Now they are implemented by computer, and typically we cannot review the form before we start because the implementation unnecessarily requires us to fill in fields before we can proceed; moreover, if we take more than 10 minutes (or whatever) filling out fields before we "send" the form back to the server, the server may time out and will discard everything because it refuses to save incomplete data. These are implementation problems that the abstraction "form"—as in paper forms—does not itself require. The implementation has added extra features, like timeouts, which make it worse for users.
- A person goes into a bank, which demands identification. The customer pulls out photo ID, but the bank says it is out of date. The customer says, "but I'm still the same person!" This joke relies for its humor on the bank's implementation bias: for them a correctly implemented ID includes the date of the ID, which of course has nothing to do with the person's identity.

The language of abstraction shows we have to take a lot of things for granted. For example, when the user's abstract model and the computer's abstract model are compared, we are already assuming that parts of the models somehow align in the first place. If my user model is that this is a word processor but the computer's model is that it is a hotel reservation system, things are not going to work out very well, whatever the models and properties.

In psychology, concepts like chunking are used in similar ways. As people acquire skill, their skill becomes more abstract and chunked, and they become less aware, if aware at all, of the implementation of their skill. Interestingly, this means that skilled users may not be very helpful when designing interactive computer systems— they have lost some of the critical awareness (for example of their errors and recovery strategies) the developers will have to explicitly implement.

Picturing Abstractions

It is helpful to draw pictures of abstractions and their relationships. To help do this, we use arrows.

The notation "X → Y" will mean that "Y is an abstraction of X." Sometimes we might informally say "X is Y" though adjusting a bit to get the English right—perhaps as in "*the* X is *a* Y"—depending on the exact sense and wording of X and Y.

Physicists are notorious for doing abstraction: cow → sphere (or in English: "consider a spherical cow…") being a famous example.

Earlier, we used abstractions to solve Kahneman's problem:

$$\text{bat} \rightarrow x$$

$$\text{ball} \rightarrow y$$

x and y were not the bat and ball, they were not even the dollar cost of the bat and ball; they were the *names* of numbers. (Did you notice the $ cost was $1.10, but that we wrote $x+y=1.1$, not $x+y=1.10$, because we also used the abstraction $1.10 → 1.1?) Note that we can put the arrow statements into reasonable English as in: "x is an abstraction of bat" or "a bat is x dollars."

The point is: abstraction is very useful, and helps us think clearly.

We also need to think clearly about what we are abstracting, and whether the abstractions we are using are the right ones. Computer programs are designed to support the user's tasks, and they are therefore abstractions of "reality." The question we are asking in this chapter is: are these abstractions the right ones, and how are we to make sure they are right or good enough for what we want?

For the purposes of this chapter the following are starting to be more useful examples:

$$\text{user} \rightarrow \text{user model}$$

Or another:

$$\text{computer} \rightarrow \text{screen display}$$

Or consider these more abstract abstractions:

$$\text{computer} \rightarrow \text{lots of bits representing its state}$$

$$\text{computer} \rightarrow \text{its on/off light}$$

We have just used four abstractions, and the four arrows above are clearly different arrows doing different things. Arrows are therefore often named to tell them apart and keep track of them. For example, the arrow in the last example above might have been called "power light." Often, the arrows are named directly, such as

$$\text{Computer} \xrightarrow{\textit{powerLight}} \text{on/off light}$$

Sometimes the name may be used with no arrow at all:

$$\textit{powerLight}(\text{computer}) \rightarrow \{\text{on, off}\}$$

This sort of mathematical notation will take us beyond the needs of this chapter.

The advantage of arrows is that we can draw *arrow diagrams*, and we can often ask insightful questions using them. For example, in some complicated arrow diagram I might wonder if I can follow arrows from A to Z, going via H, then can I get to Z if I go via M? If I go via M will it be the same as going via H? These sorts of questions make sense because abstraction arrows have an important property: if $A \rightarrow B$ and $B \rightarrow C$, then $A \rightarrow C$; or, in words, if C is an abstraction of B and B is an abstraction of A, then C is also an abstraction of A too (this property of $\rightarrow$ is called *transitivity*). Later, I will give an example how I forgot an arrow through the user's eye.

The first diagram we need for talking about critical system design is shown in Figure 1.1. Notice how we have allowed arrows to point in any convenient direction to make the diagram clearer. We also drew circles around words: the circles don't mean anything in themselves but make the diagram neater, and occasionally make it clear that several words in the circles are all about one thing.

Notice that the circles at the top are the only "real" things in this diagram; everything else is an abstraction. Obviously the diagram *itself* is an abstraction, and it does not have enough detail yet to talk about the sorts of critical problems we mentioned at the start of this chapter. An important point is that we could start introducing "implementation bias" in many ways; if we were anthropologists, for instance, we would be interested in different details than we are here when we are engineers. Here, I want to help talk about what manufacturers, specifically programmers, can do to improve the user experience.

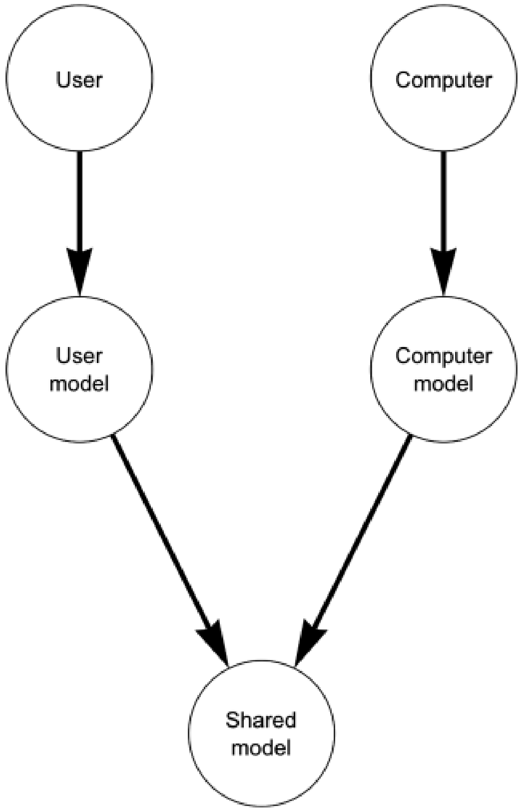


Figure 1.1 Simple abstractions using arrows. For example (top right): computer → computer model means that the computer model is an abstraction of the computer. In turn, the shared model (bottom) is an abstraction of both the user model and the computer model.

The computer model is something the user sees (for instance through the computer's display), and it is also something the user can change by interacting with the computer.

Figure 1.2, next, says the user performs actions (we are not saying how because it is an abstraction), and the computer interprets them (again, we are not saying how). The result is something that is displayed.

This of course means that "display" does not mean the colors of 700×1000 pixels, because users do not think like that. If we wanted to make this additional abstraction clearer, we could draw the diagram in Figure 1.3 instead.

This is more complicated, and not necessarily more helpful. It would be much easier is to define "computer model" to be "computer model → actual pixels" and hence abstract away from this particular implementation detail.

Computers don't just happen; they are programmed by humans. Figure 1.4 is the penultimate diagram before we start using the approach.

How the user creates the input in Figure 1.4 has been abstracted away—we could have had an arrow from the user to the input, but we have abstracted it away (this chapter isn't talking about how users work out what to do, so we have abstracted that problem).

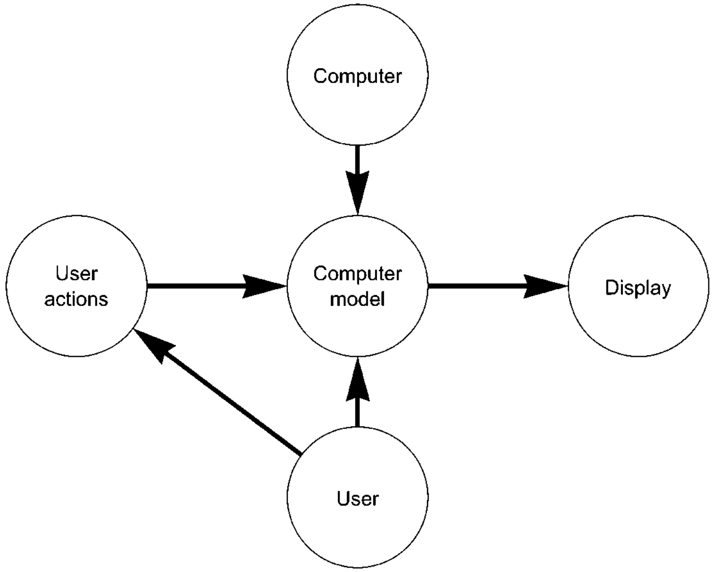


Figure 1.2 The display is an abstraction of the computer model: it shows some of the things the computer can do. User actions tell us some things about the user, so they are an abstraction of the user. Since for any arrows a→b→c we know a→c, then we can draw in an arrow from the user to the computer model as well.

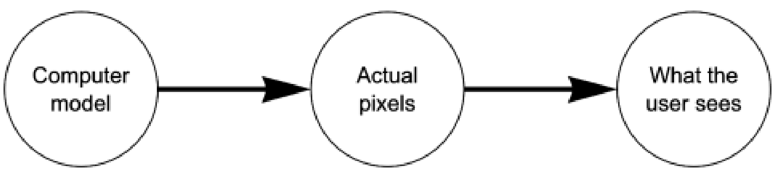


Figure 1.3 Making "display" clearer.

The two arrows on the right in Figure 1.4 have ended up in the same place (namely the output), so the diagram is saying that those objects are the same.

We can easily remember that we have to have programmers and users, so we can abstract them away to obtain a very simple diagram (in our case, we only have one of each, but in more complex systems, it might be premature to abstract them, as their social network and information flow may be critical). Figure 1.5 shows how.

Figure 1.5 again shows a single object for output. This means that whether you start with the user model or the computer model, you can abstract to the same thing. At least, that is what is *supposed* to happen if we believe the diagram (the diagram is said to *commute*—it doesn't matter which way you go around it, you end up with the same things in the same place). To make this clearer, the next diagram, Figure 1.6, is the same "up to naming" where I have replaced the names with numbers. What Figure 1.6 shows, or rather, what "it claims," is that if you start at 1 and follow the abstraction route $1 \rightarrow 2 \rightarrow 3 \rightarrow 4$ you end up with the same thing as if you go directly $1 \rightarrow 4$, because (as is obvious) you have ended up on the same thing, namely 4.

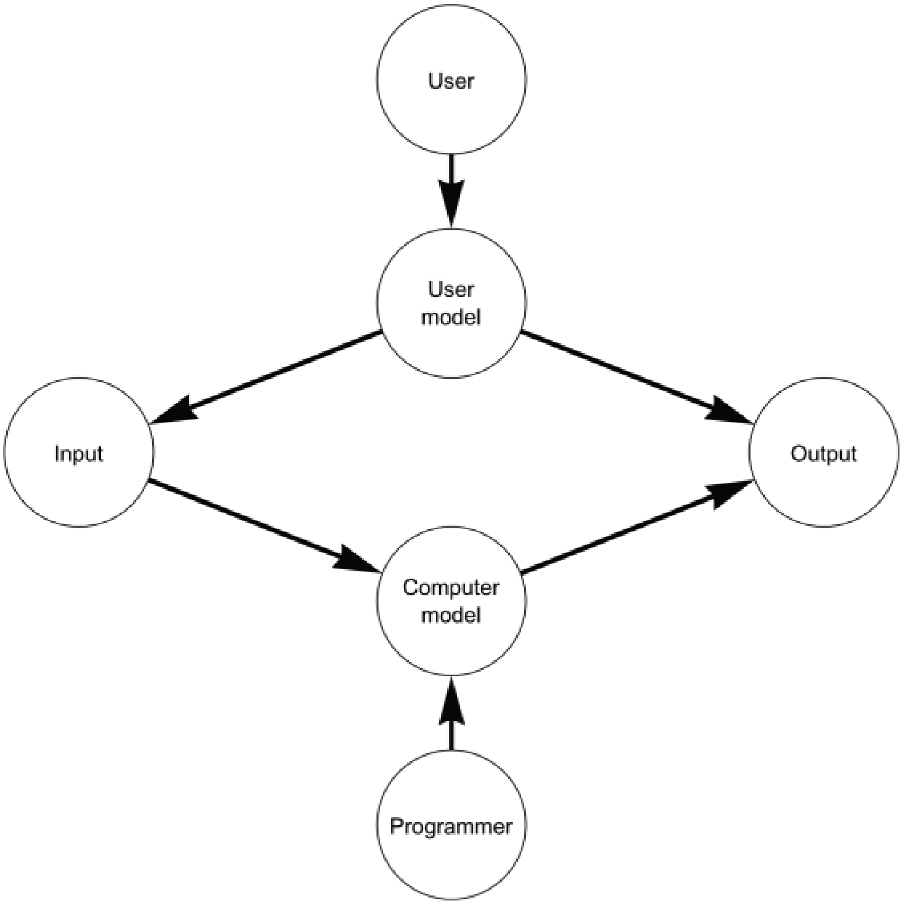


Figure 1.4 Drawing HCI abstractions, here finding a diagram with a nice symmetry between the user and the programmer.

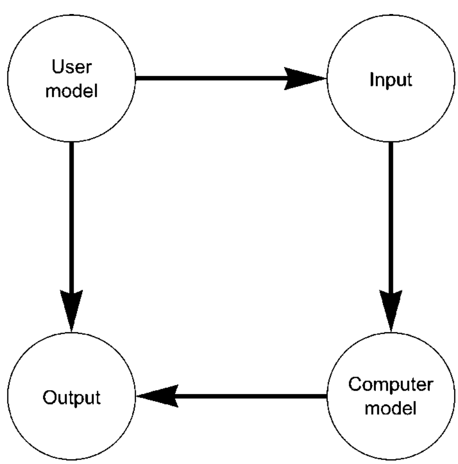


Figure 1.5 Basically Figure 1.4, but with user and programmer abstracted away—and out of sight.

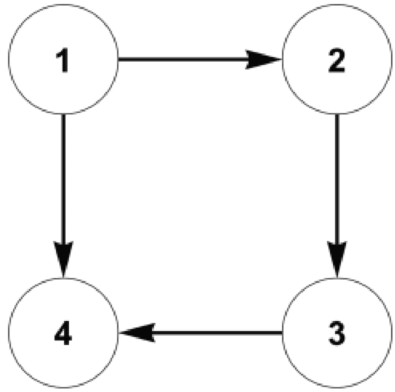


Figure 1.6 Figure 1.5 with numbers instead of names. We have abstracted away from concrete concepts to illustrate how arrows work, even when we know nothing about the ideas the numbers are representing.

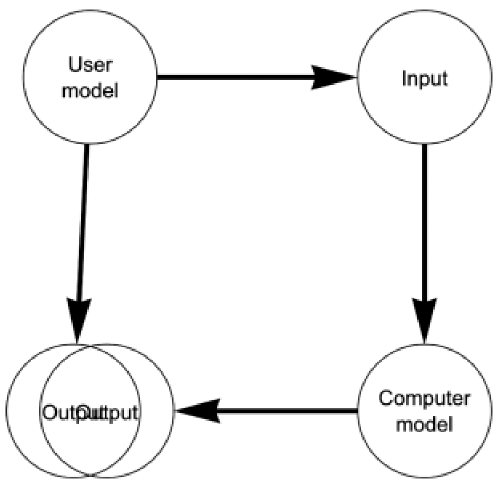


Figure 1.7 One way to draw a diagram that does not quite work. The user model abstracts to one output, and the computer model to something that is not quite the same.

In reality, these diagrams are idealizations. Figure 1.7, shows what all too often happens.

Figure 1.7 is a mess; what the computer does is not what the user thinks it does. In a conventional diagram, the circles should not overlap like this (more arrows are needed to partition the abstractions), but I like the way the output has turned into a Venn diagram—the overlapping (but not coincident) circles show that there is an abstraction that derives from *both* the user model and from the computer model, and there are things that derive from them *separately* that do not overlap.

For comparison, it is interesting to draw the diagram for one of the abstractions we used very successfully earlier in this chapter, shown in Figure 1.8.

This diagram works (it "commutes")—the right-pointing arrows are "mathematical abstraction" and the downward-pointing arrows are "solving the problem" abstractions. To solve the problem, we agree that y is (or will be!) the answer.

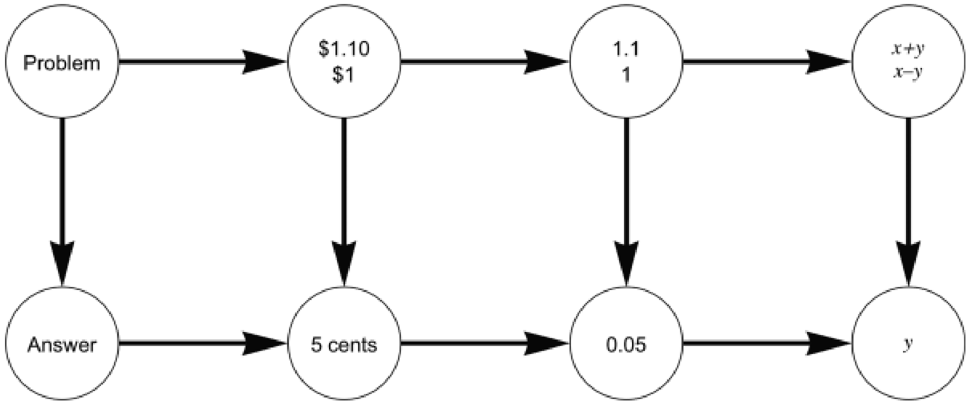


Figure 1.8 Kahneman's problem drawn as a commuting diagram.

The important point—commutativity—is that however you get from the problem on the top left to the bottom right-hand y, it is the *same* y.

As always, we can read the abstraction arrows backwards using "is," as in y is 0.05; 0.05 is 5 cents; and, finally, 5 cents is the answer.

Earlier, we casually took it for granted that we could use the abstraction \$1.10 → 1.1, then use arithmetic and algebra to get back to a correct dollar value—and the diagram above indicates that this everyday thinking is reliable, and indeed it helped us solve Kahneman's problem. (Mathematicians might go even further, and draw diagrams for + and – and so on; the abstractions and their properties work all the way down, helping get rid of distracting detail.)

We do this or things like this problem so often that all this work on abstraction seems completely obvious; indeed, we often think that \$1.10 = 1.10 = 1.1 always, *but it does not.*

What happens if a computer thinks 1.10 is 1.1 but you think they are different? For example, suppose you want to enter 1.15 into a computer but accidentally you type 1.10, so you then hit delete, to get 1.1, then you hit 5, to get 1.15. Unfortunately, your computer thinks 1.10 = 1.1, so hitting delete gets 1., then you hit 5 and you've ended up with 1.5, not 1.15. So, sometimes "obvious" abstractions are wrong! Indeed, Thimbleby and Cairns (2010) shows that you can save lives by getting them right.

In summary, we can draw diagrams of abstractions. There are many ways to do this, particularly for complex problems like HCI. For well-defined problems like Kahneman's the diagrams are simpler and much easier to agree on. Part of the benefit of abstraction diagrams is exploring and working out the best way to represent a problem.

Case Study 1: Numbers

The arrow diagrams are a notation to describe any problem but if you are not familiar with them they may seem far too arbitrary to be very useful.

For a first case study, we will use numbers—numbers are interesting because:

- what we can *see* and interact with (what is on the display, what the buttons do), and
- what they *are* (numbers for calculators, map coordinates, drug doses, money)

are different, and they are different in rather interesting ways. Furthermore, we can find ways in which programmers get this wrong, with bad consequences for users.

How are numbers interesting? Ten is equal to 10. How can that be? “T,” followed “e” followed by “n” is an English word, but “1” followed by “0” is an Arabic numeral. We have learned that the word “ten” means the number we call “ten,” and we have learned that the Arabic numeral “10” means the number we write “10” and which we also call ten, and we learned this so well at an early age it seems like they *are* the same. In fact, if we drew the abstract diagrams for numbers, we would be able to see nice patterns with arrows connecting “ten plus one is eleven” and “$10+1=11$” and so on. When we say “count” we are in fact saying “what is the numerical abstraction of that?”So •••••••••• → 10, when the arrow means “counting.”

It is useful to remember that *for centuries* we were quite sure that X meant ten, and XI eleven, IX nine, and so on. Arabic numerals are not obvious, and they are no more the “right” way to describe numbers than Roman numerals or binary numerals are.

The problem is that the arrow diagrams can become complicated and they are very easy to get wrong. More generally, it is very easy to build an interactive computer system with a broken model and not realize it. Even with something as simple and familiar as numbers.

Take a calculator implemented on a mobile phone platform as a simple example. Now, many calculator apps will differ and it is certainly another question why they differ for such simple (and one would have thought standard) things, as design variation really doesn’t help users. But let us suppose, on this one, to enter a number, the user presses the button C (“Clear”) a couple of times. The user presses it several times because they have learned in the past that this is wise. Indeed, on the second press the button turns into AC (“All Clear”?) So we may as well press it again, now it is something else.

Now, to enter a number (say 42) the user presses the keys 4 then 2. The calculator will display 42. If the user wants to delete a digit, perhaps entered in error, swiping towards the right over the number will delete the rightmost digit. So far so good. Many other calculators work like this, although perhaps using a dedicated DELETE key instead.

So far, the arrows in our earlier user model diagrams all match up (they *commute*). But something interesting is going on that has not yet surfaced to break our abstraction. We think that swiping deletes a digit, but the calculator did not do that, as it had already turned the digits 4 then 2 into a number 42. These are usually thought to be “the same thing” but one is two digits, and the other is a value (which *coincidentally* we write in conventional Arabic notation using the same two digits).

To implement the “delete,” the calculator basically divided 42 by 10 and threw away the remainder. It did a numerical calculation to turn 42 into 4, whereas the user thought “just delete the digit 2.” You can see something slightly complicated is going on behind the scenes if you delete both the digits 4 and 2 with a further swipe. The calculator does not display nothing (as if it had deleted 4 and 2); it displays 0, which is a number you did not enter. Well, that’s because 0 means nothing and deleting 4 and 2 leaves nothing—except nothing is displayed as the symbol 0. Technically, nothing → 0.

As a happy coincidence when 0 is multiplied by 10, and the next keystroke added to it, it will display that keystroke correctly. For example, if the display is 0, and the user hits 4, then $0 \times 10 + 4 = 4$, and when the user keys 2 after this, then $4 \times 10 + 2 = 42$ and the calculator displays 42. So if the display showed 0 and the user keyed 4 then 2, then the display will show 42, and it does this straightforwardly. This is what we expect of course. (It gets a bit more complicated after the decimal point is pressed; we'll worry about that later.)

Now let's try something marginally more complicated. Let the user enter –42. Let us imagine that the app lets a user enter −42 in many ways. There is a ± key. You can enter ±42, 4 ± 2, 42 ± (we already have a user model alignment problem: I want to enter a minus number, but I only have a ± button). Usually we ignore, abstract away, from these "little" details and press on.

Let's say we are entering −42 and we key ±, 4, then 2. Then we accidentally key another ±. That's OK, because we have a delete: so we swipe and we discover that the number entered is not –42 but 4. What happened? We keyed ±42 and at that stage the calculator showed −42. We then keyed the accidental ± and the display changed to 42. We then swiped, and swiping deleted the last digit in the display—namely 2—not the last key we pressed—namely the second ± .

The display does not show what the user keyed; it displays an abstraction of it. Unfortunately, most users will take "delete" to mean "delete what I just did," not do something complicated with what I can see. Worse, the developers of the app have decided that it is easier to ignore the "error" of the user keying ± at any time—rather than requiring it to be first—so they have abstracted away from *when* the ± key is pressed. When it comes to deleting it, the calculator has no idea. Although the first time ± is pressed it will be displayed as −, further presses won't be displayed as−− −, etc. What the ± key does is change the state of the number, not the way it is represented. So the delete can't possibly work correctly with the ± key.

We can see this conceptual error in the minds of the developers if we now consider how *three* presses of the delete key would be handled according to their model. If users enter, say, −42 (they key ±42 but it is displayed as −42) and they delete three times, things will go really wrong. The first delete changes the display to −4, the next delete goes to −0, and the *next* delete goes on to display − NaN.

NaN means "not a number," which means that the application to simulate deleting keystrokes has failed. I had explained a similar problem with a specific device and got it published (Thimbleby, 2015). The manufacturer corrected this bug *but introduced new ones* which we do not need to parade here. The point is that user interfaces are difficult to design and implement correctly—even simple, well understood things like calculators—and a leading company with leading programmers finds it difficult, even when the problems are pointed out.

Summarizing, then: the user keys digits and presses delete—the user's model revolves around numbers as concrete strings, which we usually call numerals. Inside the program, however, the programmer apparently thought numerals are numbers. It is a lot easier to do arithmetic on numbers than edit strings (numerals)—besides (I suppose the programmer thought) the calculator does display a number doesn't it? Unfortunately, at the boundaries, numerals and numbers do not behave in the same way, and delete (or swipe) "breaks the abstraction."

Had the programmer thought their job was to implement the user's model, then the display would show the numeral, it could be edited without problem, and when the user presses + or divide or whatever, the calculator would do its arithmetical magic as instructed. The user interface would have been easier to use.

In abstraction terms, the calculator implements a computer model that says "anything the user keys" → number. It is a calculator whose whole purpose is to show numbers! But this model fails when what the user keys is *not* a number, for instance when they swipe "delete" or press several decimal points. Unfortunately the calculator implements things using numbers not strings, so no amount of tinkering will get the user model right.

Ignoring multiple decimal points also has another problem: if users press the decimal more than once, they are making a mistake that the calculator ignores. Ignoring user error means users will not know they have made errors, and they will likely carry on obliviously. Our experiments (Thimbleby & Cairns, 2010; Thimbleby, Oladimeji, & Cairns, 2015) showed that detecting such errors and blocking them could halve the final error rate.

Imagine this calculator app with the same underlying concept of "numbers are not strings," and it can run both in landscape and portrait mode on the same handheld device. If the handheld device is turned to landscape, the user can enter very long numbers. Turn the handheld device to portrait, and a long number like 56585600000000 may be displayed as 5.65856e13 (the "e13" means "times 10^{13}"). Yet you didn't key in the decimal point shown, and there isn't even an "e" key on the calculator.The last digit you keyed was "9," and one isn't even shown in portrait. So how does delete work now? The display will turn to 5.65856e12.

Again, we see that the calculator implements delete as a numerical operation rather than a string or user interface operation. Underneath, as implemented, "e" means "times ten to the power of," so 5.65856e13 means 56585555975599 (but we can't display that when the handheld device is in portrait as it is not wide enough to do so) and the delete divides by ten, so e13 becomes e12 rather than e1. Arguably, it probably *has* deleted the last digit of 56585555975599, but the user can't see that it has.

No doubt users would be frustrated with this interface. But they would mostly be frustrations of the first kind—perhaps even unnoticed—and if used in safety-critical situations, no doubt users would "sanity check" the results. Or would they? It is likely that the wrong conceptual model of how numbers are represented has made its use a little more unsafe than it need have been. But we will never know by how much.

Consider, then, the design of a calculator that has been designed by people who know that calculations can be critical, and that a record may be required either to double check and help correct mistakes, or to prove that a correct calculation actually happened. Errors here, if noticed, will be of the second kind: annoying, and time wasting. But if the paper log were to be used as evidence or a record of some actual transactions that were supposed to happen then we start to enter into frustrations of the third kind—the safety-critical ones.

Let us suppose that this calculator is designed for financial use so it always displays a decimal point, whether or not one has been keyed. Now, the user accidentally types in a number and a point "88." and realizes the error: two decimal points! The user press the "delete" key. But delete cannot delete it, because it "has" to be there. Therefore delete does something else, in this case deleting the digit keyed before either of the two decimal points. It does this without saying so, and probably beneath the user's awareness. The user has now got "8" in the register.

It would have been far easier to avoid these problems than to describe them! However arcane these problems seem, remember that they will eventually catch people out, and the small frustrations may well, in some situations, turn into big ones.

With mass market devices the "small" frustrations are multiplied by the enormous user base the devices have. The probability of error per hour p may be small, but multiplied by the number of user-hours N, pN is the expectation of error (somewhere in the world) per hour, and it is not small enough to ignore.

Ironically, programmers work to tight schedules. So if they turn out designs quickly, they save their employers money (and perhaps earn their bonuses). But the seconds that programmers save in ignoring details of user interface design, if they get a large market share, will add up to lifetimes for users. This is technical debt.

Case Study 2: Medical Device Directive 93/42/EEC

Let us consider a medical app for use on a tablet or handheld device in a medical context. Let us also notice that is "CE marked," which means that it is approved in the European Union for such medical use (it complies with the relevant conditions in the Medical Devices Directive 93/42/EEC). In this case, the app's purpose is to calculate the appropriate fluid dose for a patient who has been burned.

The app displays an outline of a human body. Using a mouse (on a PC) or finger (on a tablet), the burned areas on the body can be sketched—let's suppose, the patient's face and left arm are burnt, equal to 13.4% of the body surface area.

The user draws, on a picture, the position of the patient's burns are and how extensive they are, enters the patient's age and weight, and the app then calculates the dose. A widely recognized formula, Parkland's formula, is used to do this.

As there is a small risk the app may be corrupted or that the tablet or phone it is running on may be damaged, a careful designer starts by making the device perform thousands of calculations to check they all get the right result from Parkland's formula (this *is* safety critical, so this is good programming practice—what is called an *integrity check* in the trade). If they don't, something has gone wrong, and the app declares itself unsafe to use.

Parkland's formula is an *invariant*, something that is (for this program) supposed to be always true or unvarying. Ideally one might wish to prove that an invariant is true, but here the point is that something may be unreliable, so the program tests the invariant with a wide range of test cases.

Unfortunately the invariant checked is only between the computer's tables of *internal* numbers, and not with anything the user *interacts* with, such as the numerals they type in. It is, in fact, possible for the user's idea of the app to be very different from the internal views the app has of itself—something our abstraction diagrams can help us think about.

For example, if the user enters a large number (perhaps accidentally—keys can autorepeat, and an ordinary number might end up being followed by lots of 0s, which then make the number very large) the number is displayed in a box that is not big enough to show the full number entered. What the users see is not what they have got. The standard does not say anything about "sanity checking" user input, so it is possible to enter unreasonable numbers (say, a patient weight larger than the weight of the Earth), and Parklands will be applied correctly to that, and thus generate a misleading instruction for treating the patient. Needless to say, such gross errors will be immediately obvious in a clinical setting but the possibility remains.

More interestingly, the invariant check does not help ensure that the numbers the user is entering are the *right* numbers. There is no requirement for a "new patient" button and no requirement for a check on the amount of time elapsed between entering readings. So the numbers may already be prepopulated by a previous patient's incompletely entered details. This means that it is possible to enter some data, put the device down, later on enter some details of a new burn patient, get Parkland's results—but the recommended dose is computed on the basis of a composite of data. Surely little would be lost and much gained by checking whether it is still the same patient.

We can imagine more subtle user interface problems too. The user draws (paints) the burns on a drawing of the patient. Rather than having two drawings for front and back displayed together, to make this easier and more accurate, the drawing can be turned over, so burns on the front and back are drawn separately, each therefore being given more screen space. There is also a numerical box to enter a percentage instead. If clinicians know that the burn area on the back is, say, 45%, they do not need to draw it as it can be entered directly as a number in this box. This means that the user may draw burns, turn the body over, then enter a percentage as a number, and all the drawn burns are deleted. The user does not see this happening as the other side of the body is hidden from view.

Of course, the percentage is (supposed to be) burn area as a percentage of the total body area, not the body area that can be seen at any moment, front or back. Do users know that their model is different from the app's? How can they find out when the app does not warn them?

From an abstraction point of view, this is the same problem that we encountered with the calculators. A user enters numbers on the app (now percentage burned body area, patient weight, etc.) and the "digits" can be either gestures drawing burns on a picture, or they can be digits forming part of a conventional Arabic number like 45. The way the app is designed, and unbeknownst to the user, these two sorts of numeral must not be mixed, even though the user interface design allows them to be.

The problems are, as before, somewhat tedious to explain. It would have been far better to prevent the problems than to spend time describing them (or, worse, experiencing them). How could they have been avoided?

One clue is in the use of invariants. Let us assume, following good practice and the need to achieve the certification, that the app tests some invariants in the program. However, checking these invariants is obviously not enough.

The correct invariant is user model = computer model, or more specifically user model = external computer model (the user probably does not care what goes on inside the computer). Instead, the app simply checked that the internal computer model = Parkland model. We could get to the right invariants from working out what the abstractions are.

The invariant the app actually checked is something like this

$$V = 400 \times m \times A$$

meaning the volume to be infused (over some hours, but that is a detail we will ignore here) is 400 times the patient's mass times their percentage burn area (plus many other details that we will simply ignore here). The app then checked this invariant for many numbers (from a table) and checked that the invariant was calculated correctly. Note that if the table itself has become corrupted then the invariants won't work, so it is a good method to help spot problems.

The invariant should have been more like $V = 400 \times m \times A \vee$ error (the $\vee$ symbol means "or"). Either the Parkland formula is calculated correctly or an error has happened. But what sort of errors might there be that need checking? Well, to start with, the user might have entered a percentage less than 0 or greater than 100, so we can start to refine error as follows error $= A < 0 \vee A > 100$, and of course it would be an error if the patient had a negative weight $m < 0$ and so on. Then there is the interesting question: what is A (or m etc.) when the app does not know it because the user hasn't specified it? The invariant needs more work: error = unknown(A) $\vee$ unknown(m) $\vee$ $A < 0 \vee \ldots$

Now we start thinking about error, we can think of more important things to check. Part of the invariant really ought to be that not only are the numbers what the user thinks they are but the numbers actually refer to the patient in front of us! So, an even better invariant is:

$$\text{error} = \text{wrongPatient} \vee \text{unknown}(A) \vee \text{unknown}(m) \vee \text{malformed}(A) \vee \text{malformed}(m) \vee A < 0 \vee \text{LOTS MORE}$$

Not only is using invariants good practice but thinking about them in detail is even better practice. Here, we found that the original invariant should be refined. Fortunately it is, in fact, easy to refine the invariant. But refining the invariant creates more design problems to solve. How does the design know we have the right patient? How does it know what a malformed number is? As we can see in the "final" invariant written above where it says "LOTS MORE" we now need to start thinking about how we can reliably know when we have all the possible errors sorted out and expressed in the invariant!

When the app checks "wrongPatient" it could put up a dialog box that says "Are you sure you have got the right patient? Yes / No?" and then the app can answer the question. Of course there are other ways of doing this, using patient barcodes, for instance, or even deciding that "wrongPatient" is too tedious to work out. Certainly invariants can have a lot of trivial things in them that can be culled—for example, checking the battery is a bit pointless for an app because with a flat battery it won't work anyway. Maybe we should just warn the user to be careful.

These are important design questions that slipped through the Medical Devices Directive 93/42/EEC. One could imagine an exhaustive set of user tests that demonstrate that the device works correctly and delivers the correct dosages according to the Directive when the inputs are correct but which do not allow for the unexpected event. As one can't expect the unexpected, the only recourse is to use mathematical reasoning during the design process.

Case Study 3: Loss of Situational Awareness

I designed and wrote an interactive program that used colors to help make the user interface clearer (Thimbleby & Gimblett, 2011). Inside, computers code colors in RGB values, as amounts of red, green, and blue. The RGB values are displayed in pixels on the screen, and they can create any color; for example, red and green together will display as yellow. It is interesting, at least to me, that colors in the programming language I was using used hexadecimal numbers, which have no "affordance" for what they represent. (CSS is another widespread example using hexadecimal coloring, although it also allows color names such as "Blue.") This issue is rather like the number/numeral distinction we considered earlier: a color and its RGB value are like numbers and numerals. It is a happy coincidence that an RGB value is a number—often written in hexadecimal x000000 rather than ordinary Arabic base 10 notation.

At one stage in designing this program I had already used red and green and I needed a new, third color, so I chose blue (I am not so stupid as to use the same color twice, so I needed a different color). So my RGB number values inside the program were x770000, x007700, and x000077.

The program worked very nicely and I was happy with it.

Then my first user tried it, and asked me immediately why I had unreadable screens, mixing red and blue! Didn't I know some people are red/blue color blind? Didn't I know red and blue is the worst combination? They are at opposite ends of the spectrum, and, particularly for people with glasses, they refract very differently and separate out making screens hard to read.

I had been unaware of it but my user model and my computer model had obviously parted company. They were not aligned.

Actually, I know all that when I think about it. What is interesting is that because programming is so hard (I tried to give a flavor of that by mentioning hexadecimal numbers) I had no spare brain capacity to think about anything other than having a different RGB value. I had got a different RGB value and the colors were certainly different (I didn't even need to check, *of course* they're different). The problem about "no spare brain capacity" (otherwise known as tunnel vision or loss of situational awareness) is that you do not even know you have insufficient capacity: you are so focused on the problem to hand (here, programming with hex RGB numbers) that you do not think about other things and you are unaware that you are not thinking about wider issues.

In abstraction terms: I had failed to think about the arrow from my RGB numbers into the user model, through the user's eye's lens and via the retina at the back of the user's eyes. I had not drawn the abstraction diagrams and checked that they commuted. Worse, it was such a simple programming decision (RGB, that's all!)—it never crossed my mind it was complicated enough to go and think about visual perception as well as just programming it.

I wish my user had been looking over my shoulder earlier to help me think more about design!

The Sharp End and the Blunt End

Users work at the "sharp end" where things happen. In HCI we have tended to focus a lot on what goes on at the sharp end, for this is where user experience happens. It is certainly important.

In healthcare, for instance, caring for and curing patients happens at the sharp end. If something goes wrong, it unravels in seconds quickly at the sharp end. The users get the blame—because they pressed the button and they caused the problem.

For example, performing tasks is demanding, so users block out distractions, losing (or at least compromising) situational awareness. Typically users are blamed for such human-factor failings. Remember how the accident investigation into the Air Inter flight ITF148 cited a “crew workload peak”?

But important things also go on at the blunt end. The designer, developer, and manufacturer spent years designing and building the system the user is using, probably building from experience with established products. The designers had an opportunity to design with care and avoid many of the problems of which the user will, at the sharp end, be *predictably* unaware—details like how the delete key works perhaps.

Unfortunately designers have tunnel vision. It is hard enough making programs work at all, let alone getting them to work well. The calculator *seems* to work; it probably does work. My RGB color choices seemed to work (and I quickly went on to worry about “more interesting” parts of the program), so I just thought they worked.

The problem is that the designer is too easily caught up in complex tasks like programming and is unable, because the programming / design task is so demanding, to step back and look properly at the abstractions. They should have been using mathematics but they were programming. It seems to work, and they can program a bit more, and it will work better surely? Unfortunately, it won’t unless the abstractions that they are implementing are right.

The conventional approach to improving usability (e.g., as Landauer, 1995, emphasizes) is to tell programmers more about users, tasks, and preferably human factors. If only programmers just understood user-centered design! Paradoxically, this makes the problem worse. Programmers have a hard enough job getting the program to work at all, let alone understanding user needs *as well*! This is itself basic human factors: the user and their task is part of the wider situation the programmer loses awareness of if they are to get their code to work. The standard solution to this problem is iterative design (e.g., ISO standard 9241), but even that has the problem that blind spots (from tunnel vision) in a design process can be perpetuated through iterative design.

The bad news, then, is that programmers have to improve how they work. The good news is that products take years to develop, and if programmers wanted to, they could program much better, easier, and safer user interfaces. We could help them with multidisciplinary teams sharing the load, assisting them to be more resilient.

Going Further

Abstract models of interactive systems seem to be rare, but there are several promising areas that are worth pursuing:

- Smith and Koppel (2013) is an excellent analysis of design problems that occur in healthcare. They point out the “misalignments” between patients, clinicians, and healthcare computer systems using similar techniques to those used in this paper. Their work shows that the abstract approach can provide very high-level insights into design of effective systems.
- Dix, Harrison, Runciman, and Thimbleby (1987) develop a PIE model, and apply the abstract ideas to sequences of input commands and their properties. It may be an old paper, but the ideas are still very relevant.
- Masci, Zhang, Jones, Curzon, and Thimbleby (2014) have demonstrated that automatic verification tools can be used to analyze human-machine interaction

design, and spot latent design anomalies before use errors and incidents happen. A wide range of problems of the sort this chapter discussed can be identified automatically: for example issues identified in commercial medical devices can be watched on YouTube (https://youtu.be/T0QmUe0bwL8).

Conclusions

Until manufacturers and programmers are informed about formal methods, and until they are motivated, things will not become either easier to use or safer. While users can be blamed, rather than the systems they use, attention will be misdirected away from design to users. Drawing abstraction diagrams may be a good way of helping think more clearly; these diagrams are much easier to discuss, change, and negotiate, and they are much, much easier to learn than full-blown formal methods.

In the 1960s, car manufacturers said "drivers have accidents" so it's not their fault things go wrong. After Ralph Nader's (1965) *Unsafe at Any Speed*, the culture changed dramatically, although not quickly. Today manufacturers say drivers have accidents *and therefore it is their duty to make cars safer.*

Cars are critical systems that are as hard to make safe as programs; the invariants cover all sorts of issues, from crumple zones to ABS brakes, skidding tires, air-bag explosives, and interacting with a complex world. Inside most computers it's usually just numbers, text, databases, and one user sitting still using office software; it ought to be easy!

I went back to that website I complained about earlier, and it said "unexpected error."

What?! They have written a program that *expected* this error, else they would not have detected it. How, then, can it be unexpected? What they really mean is, "I think there is some sort of problem here I don't know how to program for, but I am far too busy writing the rest of the program to help sort it out for you."

This chapter has presented a framework for thinking about user interface design, and we used some real examples of interactive systems to show how critical user interface design can be, and how outcomes (such as aircraft and patient safety) can depend on small design decisions that perhaps did not receive the attention they deserved. But our analysis shows that rigorous thinking about interactive systems is very hard, and it is very hard to get right. Perhaps a larger point is that iterative design and continual improvement (as recommended in the ISO standard 9241) should be taken seriously—as well as the requirement to obtain user evaluations. None of the examples, especially the company's website, provide any way to receive feedback from users, so the manufacturers cannot learn directly from problems users experience. This is a great shame.

As a user I am not interested in the error, unexpected or otherwise. I want to carry on using my computer and I want it to closely fit my model of what it should do; I am sort-of happy to go on training courses so my model better fits the computer's but never am I happy when the computer has ignored me, especially when I think the programmers have spent years, probably, at the blunt end persistently ignoring me and other users and their tasks. Let's hope, then, that programmers more often think they are building critical systems their users will be able to depend on, because they thought abstractly and hard enough about the properties that will ensure the safe and successful outcomes the user wants. Human factors experts may need to remind them as they become lost in coding, but the users' critical work and the impact it will have on everyone around them at the sharp end is more important.

Acknowledgements

Michael Harrison and Daniel Jackson gave me many great ideas.

References

Allman, E. (2012). Managing technical debt. *Communications of the ACM*, *55*(5), 50–55. doi:10.1145/2160718.2160733

Aviation Safety Network (1992). *Accident description: Air Inter Flight 148*. Retrieved from www.aviation-safety.net/database/record.php?id=19920120-0

Dix, A., Harrison, M. D., Runciman, C., & Thimbleby, H. (1987). Interaction models and the principled design of interactive systems. In H. Nichols, & D. S. Simpson (Eds.), *Proceedings of the European Software Engineering Conference* (pp. 127–135). doi 10.1007/BFb0022105

Kahneman, D. (2012). *Thinking, fast and slow*. New York, NY: Farrar, Straus, & Giroux.

Landauer, K. T. (1995). *The trouble with computers*. Boston, MA: MIT Press.

Masci, P., Zhang, Y., Jones, P., Curzon, P., & Thimbleby, H. (2014). Formal verification of medical device user interfaces using PVS. *Fundamental Approaches to Software Engineering*, *8411*, 200–214.

Miller, C. (2013). *Security considerations in managing COTS software*. Retrieved from https://www.us-cert.gov/bsi/articles/best-practices/legacy-systems/security-considerations-in-managing-cots-software

Nader, R. (1965). *Unsafe at any speed*. New York, NY: Grossman.

Smith, S. W., & Koppel, R. (2013). Healthcare information technology's relativity problems: A typology of how patients' physical reality, clinicians' mental models, and healthcare information technology differ. *Journal of the American Medical Informatics Association*, *21*(1), 117–131. doi:10.1136/amiajnl-2012-001419

Thimbleby, H. (2015). Safer user interfaces: A case study in improving number entry. *IEEE Transactions on Software Engineering*, *41*(7): 711–729. doi 10.1109/TSE.2014.2383396

Thimbleby, H., & Cairns, P. (2010). Reducing number entry errors: Solving a widespread, serious problem. *Journal of the Royal Society Interface*, *7*(51): 1429–1439. doi:10.1098/rsif.2010.0112

Thimbleby, H., & Gimblett, A. (2011). Dependable keyed data entry for interactive systems. *Electronic Communications of the EASST*, *45*, 1/16–16/16. doi:10.1145/1996461.1996497

Thimbleby, H., Oladimeji, P., & Cairns, P. (2015). Unreliable numbers: Error and harm induced by bad design can be reduced by better design. *Journal of the Royal Society Interface*, *12*(110). doi:10.1098/rsif.2015.0685

2

Semiotics and Human-Computer Interaction

Clarisse Sieckenius de Souza

Introduction

This chapter presents a broad picture of semiotics and how it can contribute to human-computer interaction (HCI). The content and references privilege clarity and cohesion of presentation, rather than coverage of existing work. The list of references comprises publications by authors with very different perspectives, which interested readers are invited to explore.

Although the definitions of semiotics may vary substantially (Nöth, 1995; Trifonas, 2015), the primary aim of a semiotician's work is to investigate signs and signification (Chandler, 2002; Eco, 1979). A sign can be defined as anything that can be used to represent something else, and signification is the process by which signs come into existence. This is not an exclusively human capacity. Animals, for example, signify the world around them, and there is controversy about whether machines can do it too (Nadin, 2007; Nake & Grabowski, 2001; Nöth, 2002). Signs are not necessarily verbal, so interdisciplinary researchers have taken HCI as the object of applied semiotics studies (Andersen, 2001; Barr, Biddle, & Noble, 2005; Benyon, 2001; de Souza, 1993; Goguen, 1999; Keeler & Denning, 1991; Mullet & Sano, 1995; Nadin, 1988a, b). A semiotic account of humans in HCI says that signs are produced by users, but not only users. Systems designers and developers, for example, engage in signification processes that are just as important (and as deserving of rigorous investigation) as the users' (Andersen, 1992; de Souza, 2005a; Kammersgaard, 1988; Nadin, 2001), and so do other stakeholders in the information and communication technology arena.

A Semiotic Vignette

The following HCI vignette shows what a semiotic perspective looks like, with computer-generated and computer-mediated signs unfolding in a piece of fiction based on true facts.[1]

[1] All interactions were functionally and interactively tested by the author in November 2015 with existing systems whose identity is not provided in order to keep the focus of narrative on technology, rather than companies.

The Wiley Handbook of Human Computer Interaction, Volume 1, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

Context

Mr. User is a retired engineer who spends much of his time on the Internet, communicating with family members and friends, watching videos, reading the news, and browsing for shopping opportunities. He just discovered that his favorite browser has an Autofill tool that can save him time and memory. He decides to explore Autofill and begins by reading online help information, which says:

> Save time with Autofill. It stores the information you type in when filling out a form and allows you to reuse it the next time you need to inform your credit card number, name and address. You can add, update, and delete information whenever you wish. Autofill is online registering and purchasing made easy!

Encouraged by this message, Mr. User engages in interaction aiming to configure the use of Autofill. The six-step storyboard in Figures 2.1, 2.2, and 2.3 shows how his experience develops. The annotations on screen sketches indicate the sequence and the logic of interaction.

Storyboard

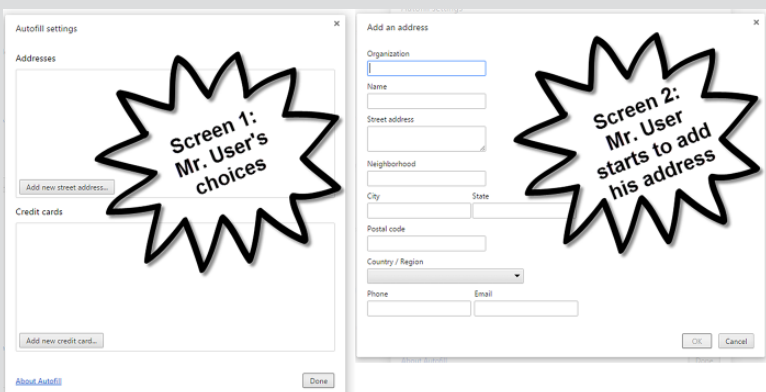


Figure 2.1 Storyboard steps 1 and 2.

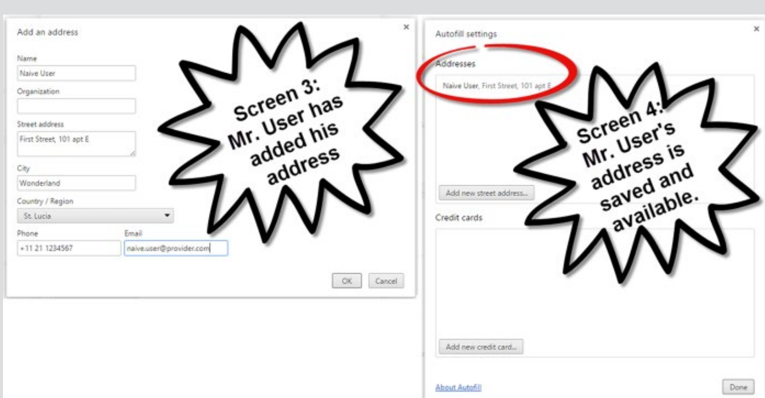


Figure 2.2 Storyboard steps 3 and 4.

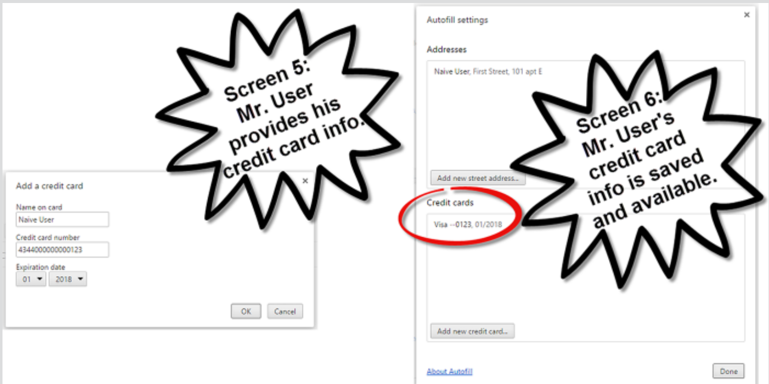


Figure 2.3 Storyboard steps 5 and 6.

The vignette illustrates how a naïve user can decide to store critically sensitive information in his local browser's database without ever questioning whether this is a risky decision. To be sure, in 2015, when Autofill was available, online theft of credit card information was a nightmare for credit-card administrators and many clients. Mr. User's behavior is not surprising, however, and semiotics can help us appreciate why, in greater depth and detail.

Human-Computer Interaction as Computer-mediated Human Communication

Except for the content of annotations, are the signs appearing in the storyboard's screens: (a) the browser's; (b) the user's; (c) both the browser's and the user's; or (d) neither the browser's nor the user's? Probably, (c) will be the most frequent answer. On Screen 5 (Figure 2.3), for example, labels like "Name on card" and "OK" have been generated by the browser, whereas "Naïve User" has been typed in by the user.

Effects of communication: Acting through interface signs

The expression "Name on card" is produced by a program, which is apparently the agent of the signification process required to communicate something to users through the system's interface. But what does it communicate? As the meaning of interface signs can be elicited by verbalizations (Andersen, 2001), in Figure 2.4 the balloons show how subsequent interaction steps might probably be verbalized by users.

On the left-hand side of Figure 2.4, the label "Name on card" is verbalized as a request (see "1"), to which Mr. User reacts by typing the name that appears on his credit card. His reaction is verbalized as: "It's N…" (the snapshot is taken before Mr. User has finished typing his entire name). On the right-hand side of Figure 2.4, "Name on card" is verbalized differently, as an assertion rather than a request (see "2"). Just like language used in communication, which can change the world and the mental state of speakers and listeners, interactive requests and assertions in HCI are also speech acts (Austin, 1962; Searle, 1969), achieved through computers.

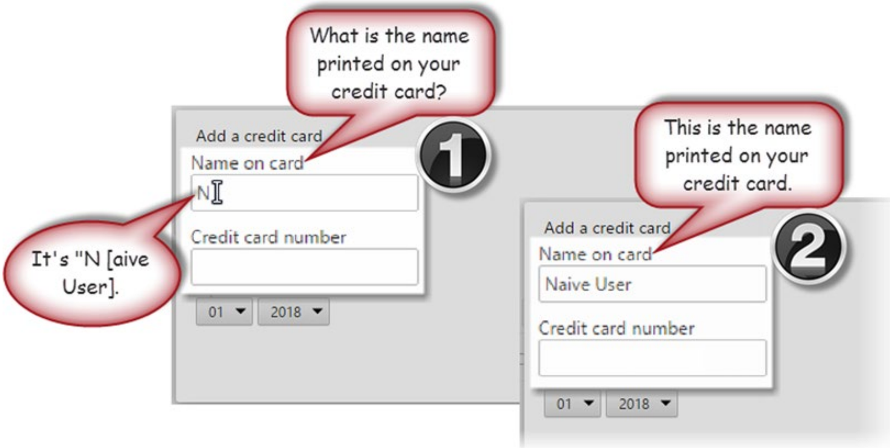


Figure 2.4 Verbalizing interface signs.

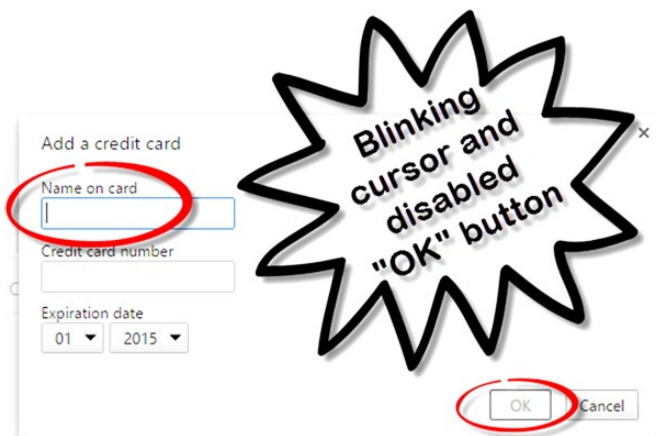


Figure 2.5 Evidence for interpreting "Name on Card" as a request.

A user's expectations, beliefs, technical knowledge and trust in developers may lead to completely different verbalizations. Starting with simpler variations, someone might justifiably verbalize "Name on card", in all turns of interaction shown in Figure 2.4, as a single assertion: "This is the name printed on your credit card." It expresses a *false* statement before the user types the appropriate name and a *true* one afterwards. In this case, as a reaction to an indirect speech act (Searle, 1975), the user corrects a wrong system state. In the previous one (Figure 2.4), the user responds to a system's request. Although these are different interpretations, as far as the system is concerned, both enable the system to move to the same expected step successfully.

In Figure 2.5 we take a closer look at how speech acts take effect in HCI and see why one interpretation is in fact more productive than the other. When Mr. User opens the dialog to add a new credit card, two nonverbal interface signs communicate something that is not strictly consistent with the truth condition interpretation. The cursor is positioned inside the "Name on card" text field and blinks (the text field's border itself has a different color from everything else on screen). Besides, the

"OK" button is disabled, preventing the user from storing credit card information if the value of this particular field is not changed. An adequate interpretation of these signs has little to do with truth conditions and much to do with requests. Moreover, as it is impossible to submit a form where the card holder's name is missing, this may be justifiably taken to mean (or signify) that the system *knows* what counts as a valid form. This conclusion would, however, be contradicted in subsequent steps. The system used in this chapter's vignette allows the user to store invalid information, like an expiration date prior to the current date, or an empty credit-card number field (which would seem to defeat the purpose of Autofill). In view of such behavior, the user might draw a final unfavorable conclusion, that Autofill cannot protect users against errors and slips.

A deeper analysis of signs and signification shows that the user might once again have reached the wrong conclusion. A savvier user, well acquainted with the risks of storing credit-card information in a browser's database, can use automatic form filling in clever ways. For example, she can fill up the form with "quasi-valid" or "quasi-complete" information. In this case, she might intentionally void the credit card number field, while filling up the cardholder's name and the expiration date correctly. The effect of this intentionally incomplete response to the system's request is that risks entailed by information theft are substantially decreased. The picture is completely changed, and Autofill now figures as a cleverly designed component, rather than a sloppy piece of software. Which one is the case?

Regardless of the answer, by going beyond the meaning of interface signs (a semantic perspective) and into the effects that they are meant to achieve (a pragmatic perspective), we gain useful insights for the design of human communication through computer proxies. For example, in Figure 2.6 the sketched interface explicitly communicates that Autofill designers intentionally allow for partial information storage, so that users can choose the level of risk they want to take.

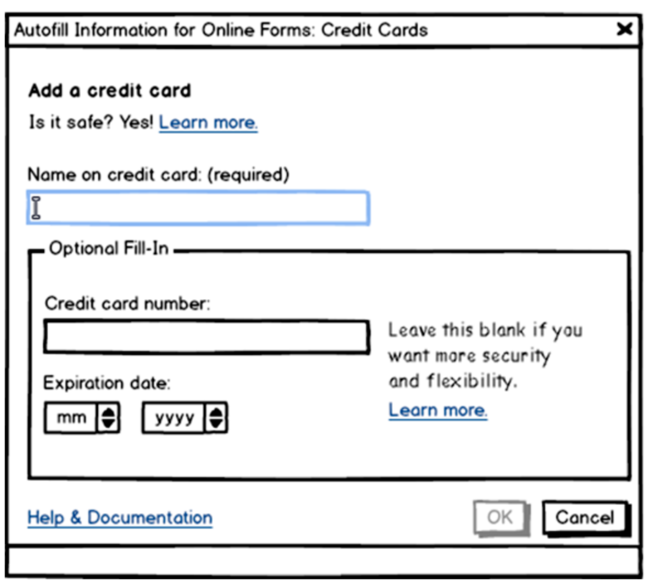


Figure 2.6 A redesigned interface for the Autofill dialog.

Who is engaged in communication?

By seeing themselves as speaking to the users through the system's interface, interaction designers can explore more often, and deeply, their own pragmatic awareness of offline communication experiences. Pragmatics can help them anticipate numerous breakdowns that have to do with shared social practices, especially if they add to their capacity of looking sharply at design *objects* that of looking at design *subjects,* which include themselves.

All semiotically informed accounts of HCI view human-computer interaction as a special case of computer-mediated human communication (Andersen, 1997; de Souza, 2005b; Nadin, 1988a). Although users are, and should be, at the center of HCI concerns, this does not come at the expense of all other parties whose messages and intent are expressed during interaction. Designers and developers are among these, but there are more players in the game like other users, systems owners, commercial companies, public and private organizations, and many more. All speak through technology deployed to benefit their own interest, and the interest of those they care for.

To illustrate the involvement of different stakeholders, we proceed with Mr. User's story, now supposing that the company that developed and distributes his browser requests access to the users' social network. Mr. User grants them the requested permission, and the company uses the fact that Mr. User has turned on his browser's Autofill function to persuade his friends to do the same.

Figure 2.7 presents a sketch of what the interface might look like for Mrs. Friend, who is part of Mr. User's social network. The text message inside the rectangle on the upper right side of the screen is a system's message telling Mrs. Friend that Mr. User has turned on Autofill. However, when she hovers the mouse over the link with his name, a balloon message pops up saying: "Hi, Mrs. Friend. I'm using Autofill."

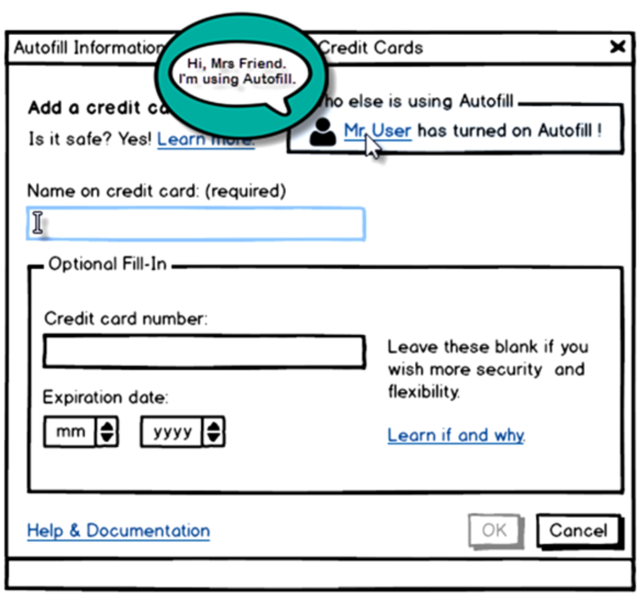


Figure 2.7 Using social network information from a user's account.

Despite the first person in the message, is it really Mr. User who is speaking? This communication is sent on Mr. User's behalf by the system, which in turn speaks for the company. He is most probably not aware of it, and therefore does not intend it to be sent.

Systems do not have intentions of their own and there is no communication without intentionality (at least presumed intentionality). Therefore, in order to achieve human-computer interaction, systems designers and developers must impart intentionality to systems. Although many decisions about the communicative intentions are derived from impersonal task and domain models, the collection of examples in this chapter shows that decisions can also derive from personal intuitions, values, and cultural practices that designers consciously or unconsciously hold. These are eventually inscribed into the systems they design (de Souza, Cerqueira, Afonso, Brandão, & Ferreira, 2016). The contribution of a semiotic perspective is to provide conceptual models, methods, and tools to support intentional design of implicit and explicit computer-mediated human communication through systems interfaces (Brejcha, 2014; Mullet & Sano, 1995; Pereira & Baranauskas, 2015).

Communicating with the mediation of an automaton

Computers play the role of medium, and also that of proxies speaking on behalf of people to other people (de Souza, 2005a, b; Fogg, 2003; Light, 2004; Nadin, 2001; Nake & Grabowski, 2001). In the latter case, a computer system executes one or more programs that contain all, and only, the rules that determine the proxy's capacity to mediate communication between the human parties involved in it. These programs can be interpreted as implementations of human communication models, through verbal and nonverbal signs (Andersen, 2001; Brennan, 1990; O'Neill, Benyon, & Turner, 2002). One of the hard challenges for HCI design is that computer programs *are* automata, everything they do is strictly determined by rules governing their response in anticipated situations that may or may not occur. This does not happen to (and is therefore not expected from) humans, whose social behavior is hardly ever the same, changing and adapting constantly to the most minuscule difference in familiar situations. The fact that humans' computer proxies are governed by specified rules, even when learning from their environment, means that their behavior is, at least in theory and unlike our own behavior, predictable (and hence *replicable*), over time and space. This opens new possibilities for social interactions whose nature and effects are, as of yet, relatively unknown (Gaines, 2015; O'Halloran, 2015). As a thought exercise, let us project some future scenario where most users will create several software proxies for themselves. The programs governing interactions with others (possibly others' proxies) will strongly signify their creators' beliefs about an important part of what they *are*, a powerful sign for investigating human *subjectivity* (Tanaka-Ishii, 2015; Turkle, 2005), as well as for investigating human *sociability* and *culture* (Brejcha, 2014, 2015; Pereira & Baranauskas, 2015; Salgado, Leitão, & de Souza, 2013).

The popularization of social communication through computer proxies can be the result of many efforts being carried out in the field of end-user development (Lieberman, Paternò, & Wulf, 2006). A larger number of users will be able to write programs, usually called scripts, to do various things for them. One thing that they can already do is to send birthday messages to friends whose birthdays are marked on their calendar. This sort of communication has the potential to outlive the senders and receivers of such messages and create a new object of study for postmortem interaction research (Pereira & Maciel, 2013).

Practical evidence of some of the consequences of our computational *persona* reaching space and time extending beyond the physical limits of our existence comes from Google. In 2013, they launched the inactive account manager (IAM), whose purpose is to help users *plan their digital afterlife* (Tuerk, 2013). Users tell Google what to do with their account information and services if they die or are no longer available. For example, an active user may have a shared agenda that automatically sends out meeting invitations to all team members of a project, on every fourth Friday of the month. If this user dies unexpectedly, this person's colleagues will keep receiving his or her invitations for monthly meetings for as long as the account is active. With IAM this can be prevented because Google users can, among other options, designate trusted individuals to take the necessary measures under specific circumstances.

The birthday card and IAM examples can be combined in script scenarios using IFTTT ("if this then that"), an easy-to-use cloud-based tool that enables users to program events using apps and Web services. One popular "recipe," as IFTTT scripts are called, sends one's friends an automatic Happy Birthday message (https://ifttt.com/recipes/114870-wish-your-friends-a-happy-birthday-on-facebook). By simply connecting Google Calendar and Facebook channels, a user can project his social interaction to time and space beyond the limits of physical life. No programming skills are required; all the users have to do is to "point and click," while IFTTT guides them through a few steps in the creation of an actual running instance of the recipe. This is already an open door to posthumous interactions, which can turn into whimsical afterlife conversations if, while living, recipients of birthday messages write their recipe to send back "thank you" notes to their friends.

Semiotic Theories and Approaches

Signs and signification processes mediated or automated by computer programs can be explored using different semiotic theories. This section presents a brief overview of concepts used in semiotically informed HCI research, and illustrates research done to date in this area. Examples are based on the same scenarios used in previous sections.

Semiotics

Semiotics has a long history in philosophy (Eco, 1986). Despite many schools and theories (Chandler, 2002; Nöth, 1995; Trifonas, 2015), most contemporary work in semiotics stems from two traditions, established by Ferdinand de Saussure (1995), in Europe, and Charles Sanders Peirce (1992, 1998), in North America.

De Saussure focused on linguistics and the role of signs in social life. *Semiology*, as he called the discipline, should investigate not only the nature of signs and the laws governing the conceptual structure and meaning of language, as an abstract system, but also the situated use of signs and sign systems by language speakers, in psychosocially determined real-world conditions (de Saussure, 1995). Peirce focused on logic, epistemology and the philosophy of science, trying to establish the conditions under which one signifies reality. *Semiotic*, as he coined the discipline, should investigate the nature of signs and sign processes through which minds continually construct and evolve the meaning of whatever becomes the object of observation, experience, reasoning or imagination (Peirce, 1992, 1998).

De Saussure and Peirce had different definitions for their object of study. The Saussurean sign is defined as a unit formed by a concept (*signified*) and a sound pattern (*signifier*). The connections between signified and signifier are established systemically by paradigmatic and syntagmatic relations. Paradigmatic relations spring from distinctions that, in some specific dimension, define classes of signs that are interchangeable (i.e. belong to the same paradigm) or not interchangeable (i.e. belong to distinct paradigms). Syntagmatic relations express the compositional principles behind complex signs, which are combinations of atomic or other complex signs from different paradigms. This was the basis for a very influential school in Linguistics and other disciplines called structuralism, whose impact has even reached the study of formal languages and computing (Chomsky, 1959; Hopcroft, Motwani & Ullman, 2013).

De Saussure's interest in how signs occur in real psychosocial situations has also led to fundamental progress in the study of natural languages. Unlike in abstractly defined conceptual systems, in the real world variability is everywhere, some of which is predictable and acceptable, while other leads to breakdowns and disruption. *Langue* and *parole* are the two Saussurean concepts used to distinguish between the abstract conceptual dimension and the situated psychosocial dimension of language, respectively. They opened the door to the study of formal linguistics, on the one hand, as well as of various other fields of language studies focusing mainly on psychology, sociology, and culture.

De Saussure emphasizes that the fundamental theoretical relation between signifier and signified is arbitrary. There is nothing in the latter that determines the shape of the former, although such determination may be occasionally verified, as in onomatopoeic words like "click." In HCI, however, although computer proxies' signs are arbitrary with respect to the form that they take in a system's interface, they are causally related to the rules that govern the system's behavior while in use. Therefore, not only the theoretical arbitrariness of the sign, but also the difference between *langue* and *parole* are blurred in a Saussurean analysis of interactive systems. A programmer's personal use of language (*parole*) becomes the system's rule (*langue*) once it is encoded in software.

Peirce, unlike de Saussure, defined the sign as a triad composed of a representation (the *representamen*), an object, and an interpretation (the *interpretant*). Interpretation is the necessary link that binds an object to some representation, and thus creates meaning. The binding can be due to habitual associations governed by personal experience or social conventions, as well as to sudden realizations or intuitions about previously unsuspected relations. In Peircean semiotics, meaning is the result of interpretive mediation that takes something—actually, anything taken as a contingent representation—to stand for something else (its object). Unless this mediation is in place, no sign can exist. Peirce's *interpretant* explains, for example, why the English word "cancel" (the *representamen*) stands for some action in the world (the object). A collective of minds establishes a conventional connection between the linguistic representation and the phenomenon or event in reference. This object can also be signified by nonlinguistic representations. The image of an interface control like ⌫ or ←, as well as the deletion of an item on a Web page list, both can be taken to *signify* the act of canceling, for somebody, in some respect or capacity. Explanations why representations signify cancelations, before or after the fact of signification, are essential for a Peircean analysis, which may involve shorter or longer chains of meaning associations, with implicit and explicit inferences.

A special feature of the mediating interpretation is that the mediator constitutes, in itself, another sign. This is to say that, among other things, higher order interpretations may occasionally bind a previous interpretation to its object if some mind performs the conscious or unconscious mental operation required for it. For example, new (higher order) interpretations (meanings) of *cancel* signs can be continually generated by computer users over time. This process, called semiosis, can take indefinitely many shapes, depending on each user's experience with systems. We may encounter surprising elements in semiosis, like the label "cancel," for instance, leading to meaningful mental images of an online purchase cancelation, which in turn lead to meaningful mental images of why this cancelation was made (for example, exasperating delays in delivering the purchased goods), which in turn lead to meaningful mental images of not trusting certain online businesses, and so on. Hence, "cancel" may signify flexibility and trust for some, while signifying nothing of the sort for others. Although theoretically unlimited in its shape and duration, semiosis is halted and resumed for practical reasons related to physical, psychological, and cultural determinations upon human experience.

To illustrate the different perspectives on HCI gained with these two theories we go back to the vignette presented in the previous section. As already mentioned, the Saussurean concepts of *langue*, *parole*, and the arbitrariness of the sign can be traced in the implementation of the interface language through which Mr. User communicates with Autofill. Screen layouts shown in the storyboard suggest certain paradigmatic and syntagmatic relations. For instance, whereas buttons and links are paradigmatically related to each other as triggers of major state changes in the interface, buttons and text areas are not. Filling up text fields typically causes more local (minor) changes of states in the interface compared to pressing a button to submit a form. Syntagmatically, the visual structure of the interface is such that the "OK" button, for instance, occupies the last *reading* position for Western users. Other interactive controls will most probably be scanned by their eyes *before* they reach the "OK" button.

Switching to a Peircean perspective, by means of mental interpretation processes, Mr. User, who is a Western user, may interpret the first Autofill dialog window presented in the vignette's storyboard as a representation of an automated process that facilitates his interaction with online forms. In this way, not only does he establish a productive signifying relation between the representation he sees on screen and the encoded behavior of the system but he also imparts a positive meaning to the interaction itself by associating the outcome of interaction with desirable future scenarios. Compared with a Saussurean account, which at first pass gives us more insights about syntax and form, a Peircean account guides us more directly into certain meaning dimensions related to the user's personal experience, expectations, beliefs, knowledge, perceptions, and so on. Moreover, the ongoing process of semiosis can elegantly account for changes in a user's interpretation of interaction signs. Thus, for example, if—hypothetically—Mr. User hears about criminal theft of credit card information carried out by scripts that intercept automatic form-filling functions executed with Autofill, his perception and interpretation will certainly change. Mr. User will probably be worried that he may fall victim to this sort of theft, and the previous meaning he assigned to Autofill will now be a sign of innocence and credulity. Note that if information theft risks can be minimized or maximized depending on how the user sets the browser's security parameters, the presence of interface signs referring to security, and theft prevention becomes crucially important to orient the user's semiosis along the desired paths. The redesign of the Autofill interface proposed in Figure 2.6 (see "Effects of communication: Acting through interface signs", above) is thus an example of how increased semiotic awareness can improve HCI design.

Peircean theory includes two other elements that, in specific respects discussed below, can bring interesting insights to interaction designers. One is a particular kind of logic reasoning called abduction, also known as hypothetical reasoning. Very briefly, abduction is an inference leading to the best explanation for a temporarily intriguing fact (Peirce, 1998). Intriguing facts have unknown meanings. According to Peirce, interpretation and abduction share the same basic principles. To illustrate how the process works, let us go back to Mrs. Friend, the main character of a previous scenario in this chapter. Suppose that since she started using Facebook 2 years ago, Mrs. Friend has received a birthday message from Mr. User through this system. Both messages made explicit reference to her little granddaughters, which Mrs. Friend thinks is especially nice of Mr. User to mention. Now, shortly before her next birthday, Mrs. Friend's son has a new baby boy and Mrs. Friend shares the news and dozens of pictures with all of her friends on Facebook, including Mr. User. He congratulates her profusely, but on her birthday Mrs. Friend receives a card from Mr. User wishing her a great celebration with "her lovely granddaughters."

What could this card mean? That Mr. User has inexplicably forgotten that she has a new grandson? That he has carelessly copied and pasted the text from last year's birthday message? Both alternatives, and many others, would be potential explanations for Mr. User's intriguing behavior. Finding the meaning of his message involves the generation and test of different hypotheses. For example, Mrs. Friend will discard the first alternative explanation because she has concrete evidence that Mr. User hasn't forgotten that she has a grandson. He posted a comment on one of the baby's photos just the day before her birthday. Although she does not know it, the other alternative explanation is also false. When he joined Facebook a couple of years before her, Mr. User created a script to send his friends a birthday card automatically. As a result, Mrs. Friend may conclude that Mr. User has been reusing old birthday cards, and that his messages are much less *personal* than they first sounded. The conclusion may become a conviction when she hears a common friend comment that he doesn't like getting always the same birthday messages from Mr. User (first piece of evidence reinforcing the "copy-and-paste" hypothesis) and when, fortuitously, Mrs. Friend's son comments that just the other day Mr. User called to ask him why he could not copy and paste text from a pdf file that he had downloaded (second piece of evidence reinforcing the "copy-and-paste" hypothesis). The effect of such meaningful but wrong inferences is that Mrs. Friend will be very disappointed with Mr. User, and what was once the source of positive feelings is now the source of negative ones. We can wonder what might happen if Mrs. Friend ever heard about IFTTT recipes, which points to the potential value of Peircean semiotic theories in the elaboration design scenarios. They can contribute to increase the number and diversify the sort of envisaged situations to which computer proxies of human beings (in particular an HCI designer's own) must respond computationally.

The other additional element of Peircean semiotics that is worth bringing up in this section is the role of mental habits in shaping continually evolving semiosic paths.

Since humans have finite resources (mental, physical, and other), although unlimited in the long run, from an individual's perspective semiosis cannot be infinite. The search for confirming evidence to support meaning generation and the generation of higher order meaning based on previous interpretations is halted when the interpreter runs out of resources like time, information, or even motivation, and patience. Mental habits play an important part in the generation of hypotheses regarding the meaning of surprising facts. Once the presence of reinforcing evidence for a hypothetical meaning saturates the need for confirmation, not only does the hypothesis gain a new status in the pool of our beliefs, but it also begins to affect the way how we interpret reality. For example, it may blind us to the observation of disconfirming evidence for certain beliefs (Magnani, 2013).

The study of signification habits and how they affect interpretation has been one of Peirce's most significant contributions to philosophy. He is the founder of pragmatism and the precursor of linguistic pragmatics (Nöth, 2011; Peirce, 1998), which we can define for the sake of simplicity as a current of thought or field of study that conceives of meaning as encompassing all the practical consequences of interpretation. For HCI, in particular, a pragmatic perspective can be extremely insightful because some of the thorniest issues in HCI design lie precisely in the practical consequences of encoding a particular mode and range of interpretations for the signs that digital proxies are able to handle in communication. For example, the practical consequences of not providing any signs of trustworthiness in the design of interaction with Autofill may be that the average user generates and sustains negative beliefs about the intentions of Autofill's developers. They, in turn, with the habit of being trusted for pieces of software that they have produced, may well forget the need to reassure and convince their users that this component can be used safely.

Despite the insights that abduction and semiosis can bring to how culture, context, and time affect users' interpretations of interface signs, the most popular use of Peircean semiotics in HCI has been his classification of signs into icons, indices, and symbols. Graphic interface designers have been using this classification since the early 1990s (Marcus, 1992; Mullet & Sano, 1995) to build interface sign catalogs and styles. In HCI, a much simpler version of Peirce's definition is used: icons are signs whose representations resemble the object to which they refer; indices are signs whose representations are contiguously (often causally) related to their object; and finally symbols are representations that refer to their object by means of established conventions. Storyboard screens in Figures 2.1, 2.2 and 2.3 show examples of the three classes of signs. A push button on a graphical interface is an *icon* of a physical push button. The labels used next to textual form fields contain words that are classified as *symbols*. Finally, the automatically updated status of the "OK" button in the form, switching from enabled to disabled and vice versa, is an *index* of the system's computation upon the content of required fields.

Precursors, pioneers, and developers of the field

Most of the existing work on semiotics and HCI stems from the pioneering work of Andersen (1992, 1997) and Nadin (1988a, 1988b). Andersen has been strongly influenced by European structuralism in linguistics, whereas Nadin is a Peircean semiotician. Bridging between de Saussure and Peirce, Eco (1979, 1984, 1986), who also investigated the semiotics of communication, has inspired other early work in HCI, such as Benyon's (1998, 2001) and de Souza's (1993). A considerable subset of pioneers' and their followers' research was additionally influenced by Winograd and Flores's (1986) seminal views on the design of computer-supported group coordination and communication systems, known as the language-action perspective (LAP). Part of it has been developed by information systems researchers, where the works of Ronald Stamper (Stamper, Liu, Hafkamp, & Ades, 2000) and Kecheng Liu (2000) stand out as references for organizational semiotics. Another community of researchers looks at information systems as *action systems,* and has been consistently exploring the consequences and possibilities of LAP. For an overview, see Aakhus, Ågerfalk, Lyytinen, & Te'eni (2014).

Language-action perspective has extended its influence in other directions, too. De Souza's semiotic engineering (2005b) is considerably different from most of her predecessors', in the sense that it is not an application of semiotic theories and methods to a new object of study, like in Andersen (1997) and Nadin (1988a), for example. It is rather the construction of a dedicated semiotic theory aiming to account solely for a single specific phenomenon: human-computer interaction. Along with other Brazilian researchers, de Souza defined semiotic engineering's object of investigation as the process of metacommunication. It is achieved by systems interfaces when activated by users for whichever purpose or reason. Metacommunication comes from the fact that systems interfaces are messages sent from systems designers and developers to systems users (top-level communication). They tell the users how, when, where, and why to exchange direct messages with the system (lower level communication) in order to achieve a certain range of effects that are compatible with the system designers' vision. In semiotic engineering, systems have been characterized as the *designers' deputy*, another term to denote the concept of a human proxy used in this chapter. Semiotic engineering has also proposed its own methods for technical and scientific studies (de Souza & Leitão, 2009), all of them centered on *communicability*, the main quality of metacommunication according to the theory.

Over the years, semiotics has been consistently attractive for researchers interested in the analysis, the design and evaluation of human-computer interaction. The variety of approaches can be illustrated by the work of Baranauskas and Bonacin (2008), Brejcha and Marcus (2013), Islam and Bouwman (2016), O'Neill (2008), and Valtolina, Barricelli, and Dittrich (2012). Moreover, a number of researchers working with culture in HCI have adopted semiotics as a theoretical foundation for their studies (Brejcha, 2015; Khaled, Barr, Fischer, Noble, & Biddle, 2006; Pereira & Baranauskas, 2015; Salgado et al., 2013). This is a particularly promising area of application for semiotic theories, given the cultural determination of signification systems and the existing body of knowledge produced by semioticians themselves when discussing culture (Danesi & Perron, 1999; Eco, 1979; Sedda, 2015). Likewise, studies on creativity and entertainment, as well as on innovative user experiences, have been and will probably continue to be influenced by semiotics (Barr, Noble, & Biddle, 2007; Danylak & Edmonds, 2005; Nake, 2005; O'Neill, 2008).

The variety and depth of semiotic research in HCI shows that semiotics can lead us beyond decisions of which images or words should be used in interface design. By framing old and new research questions in considerably different ways it can bring innovative perspectives and technological developments to all sorts of computer-mediated human communication.

Concluding Remarks

The views presented in this chapter underline the role of all humans involved in software development and specifically illustrate how the meanings encoded by designers and developers on the systems that they create can reach far in affecting not only individual users but society and culture. In this light, earlier work about contrasting the views of computers as the source or computers as medium (Reeves & Nass, 1996; Sundar & Nass, 2000) meets new alternatives. Semiotic approaches can reconcile the two opposing views, and even meet the work of B. J. Fogg (2003) on computers as persuasive technologies. However, there is yet much more to learn and do. One of the most intriguing directions in which semiotically inspired research can go is foreshadowed by studies about digital legacy (Maciel & Pereira, 2013; Prates, Rosson, & de Souza, 2015a, b). When coupled with progress made in the area of end-user development (Paternò & Wulf, 2017), it brings about thought-provoking scenarios with technology outliving users (and acting on their behalf) for an indefinitely long period of time. These can, in turn, meet theoretical work in semiotics about computers as semiotic machines (Nadin, 2007; Nöth, 2002), providing solid grounds for good interdisciplinary research.

Further semiotic investigations of computer proxies for human beings can also contribute to other areas of study such as psychology, sociology, communication studies, pragmatics, and possibly others, in addition to computer science itself. Once we begin to probe the meaning of programs from a semiotic point of view, programs become patterned signs, whose meanings can be associated with programming languages, paradigms, and styles. Programming thus becomes part of a broader cultural production process (Andersen, 1997; de Souza et al., 2016; Floyd, 1992; Goguen, 1999; Khaled, Barr, Noble, & Biddle, 2004; Noble & Biddle, 2002; Tanaka-Ishii, 2010), which can even contribute to redefine the positioning of computer science relative to other disciplines and fields. The complexity and the extent of technology's impact in contemporary life calls for theoretical foundations and perspectives that can bring together a wide variety of topics that, so far, have been studied in isolation. Andersen (2001) believes that "Semiotics is the Mathematics of the Humanities," an overarching theory that can have a powerful unifying effect in HCI as well as in other areas of computer science and information technology. It is thus worth mentioning, in closing this chapter, that in his book titled *Living with Complexity*, Norman (2010) has an entire chapter dedicated to semiotics. This may be an important sign of how different traditions in HCI studies can begin to talk to each other and bring up exciting new insights to the field.

References

Aakhus, M., Ågerfalk, P. J., Lyytinen, K., & Te'eni, D. (2014). Symbolic action research in information systems. *MIS Quarterly*, *38*(4), 1187–1200.

Andersen, P. B. (1992). Computer semiotics. *Scandinavian Journal of Information Systems*, *4*(1), 1–28.

Andersen, P. B. (1997). *A theory of computer semiotics: Semiotic approaches to construction and assessment of computer systems* (2nd ed.). Cambridge: Cambridge University Press.

Andersen, P. B. (2001). What semiotics can and cannot do for HCI. *Knowledge-Based Systems*, *14*, 419–424.

Austin, J. L. (1962). *How to do things with words.* Oxford: Clarendon Press.

Baranauskas, M. C., & Bonacin, R. (2008). Design—indicating through signs. *Design Issues*, *24*(3), 30–45.

Barr, P., Biddle, R., & Noble, J. (2005). A semiotic model of user-interface metaphor. In K. Liu (Ed.), *Virtual, distributed and flexible organisations: Studies in organisational semiotics* (pp. 189–215). Dordrecht: Springer Netherlands.

Barr, P., Noble, J., & Biddle, R. (2007). Video game values: Human–computer interaction and games. *Interacting with Computers*, *19*(2), 180–195.

Benyon, D. (1998). *Beyond navigation as metaphor.* In C. Nikolaou, & C. Stephanidis (Eds.), *Proceedings of ECDL' 98, 1513* (pp. 705–716). Heidelberg: Springer.

Benyon, D. (2001). The new HCI? Navigation of information space. *Knowledge-Based Systems*, *14*, 425–430.

Brejcha J. (2014) Ideologies in HCI: A semiotic perspective. In A. Marcus (Ed.), *Design, user experience, and usability. Theories, methods, and tools for designing the user experience.* Cham: Springer.

Brejcha, J. (2015). *Cross-cultural human-computer interaction and user experience design: A semiotic perspective.* Boca Raton, FL: CRC Press.

Brejcha, J., & Marcus, A. (2013). Semiotics of interaction: Towards a UI alphabet. In Kurosu M. (Ed.), *Human-computer interaction. Human-centred design approaches, methods, tools, and environments.* Berlin: Springer.

Brennan, S. E. (1990). Conversation as direct manipulation: An iconoclastic view. In B. Laurel, & J. S. Mountford (Eds.), *The art of human-computer interface design* (pp. 393–404). Reading, MA: Addison-Wesley.

Chandler, D. (2002). *Semiotics: The basics.* London: Routledge.

Chomsky, N. (1959). On certain formal properties of grammars. *Information and Control*, *2*(2), 137–167.

Danesi, M., & Perron, P. (1999). *Analyzing cultures: An introduction and handbook.* Indiana University Press.

Danylak, R., & Edmonds, E. (2005). The interactive game: Origins and effects. In *Proceedings of the Second Australasian Conference on Interactive Entertainment* (pp. 65–69). Sydney: Creative and Cognition Studios Press.

de Saussure, F. (1995). *Cours de linguistique generale.* Paris: Payot.

de Souza, C. S. (1993). The semiotic engineering of user interface languages. *International Journal of Man-Machine Studies*, *39*(5), 753–773.

de Souza, C. S. (2005a). Semiotic engineering: Bringing designers and users together at interaction time. *Interacting with Computers*, *17*(3), 317–341.

de Souza, C. S. (2005b). *The semiotic engineering of human-computer interaction.* Cambridge, MA: MIT Press.

de Souza, C. S., Cerqueira, R. F., Afonso, L. M., Brandão, R. R., & Ferreira, J. S. (2016). *Software developers as users. Semiotic investigations in human-centered software development* (1st ed.). New York, NY: Springer International Publishing.

de Souza, C. S., & Leitão, C. F. (2009). *Semiotic engineering methods for scientific research in HCI* (Vol. *2*). San Rafael, CA: Morgan & Claypool.

Eco, U. (1979). *A theory of semiotics.* Bloomington, IN: Indiana University Press.

Eco, U. (1984). *The role of the reader: Explorations in the semiotics of texts.* Bloomington, IN: Indiana University Press.

Eco, U. (1986). *Semiotics and the philosophy of language.* Bloomington, IN: Indiana University Press.

Floyd, C. (1992). Software development as reality construction. In C. Floyd, H. Zullighoven, R. Budde, & R. Keil-Slawik (Eds.), *Software development and reality construction* (pp. 86–100). Berlin, Springer.

Fogg, B. J. (2003). *Persuasive technology: Using computers to change what we think and do.* Boston, MA: Morgan Kaufmann Publishers.

Gaines, E. (2015). Understanding the codes and assumptions of new media. In P. P. Trifonas (Ed.), *International handbook of semiotics* (pp. 369–376). Dordrecht: Springer Netherlands.

Goguen, J. (1999). An Introduction to Algebraic Semiotics, with Application to User Interface Design. In C. Nehaniv (Ed.), *Computation for metaphors, analogy, and agents* (pp. 242–291). Heidelberg: Springer.

Hopcroft, J., Motwani, R., & Ullman, J. (2013). *Introduction to automata theory, languages, and computation* (3rd ed.). Harlow: Pearson.

Islam, M. N., & Bouwman, H. (2016). Towards user–intuitive web interface sign design and evaluation: A semiotic framework. *International Journal of Human-Computer Studies*, *86*, 121–137.

Kammersgaard, J. (1988). Four different perspectives on human-computer interaction. *International Journal of Man-Machine Studies*, *28*(4), 343–362.

Keeler, M. A., & Denning, S. M. (1991). The challenge of interface design for communication theory: from interaction metaphor to contexts of discovery. *Interacting with Computers*, *3*(3), 283–301.

Khaled, R., Barr, P., Fischer, R., Noble, J., & Biddle, R. (2006). Factoring culture into the design of a persuasive game. In *Proceedings of OZCHI '06 The 18th Australia conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments* (pp. 213–220). New York, NY: ACM.

Khaled, R., Barr, P., Noble, J., & Biddle, R. (2004). System metaphor in "extreme programming": A semiotic approach. In *Proceedings of the International Workshop on Organisational Semiotics* (pp. 152–172). Setúbal. Portugal: INSTICC Press.

Light, A. (2004). Designing to persuade: the use of emotion in networked media. *Interacting with Computers*, *16*(4), 729–738.

Liu, K. (2000). *Semiotics in information systems engineering.* Cambridge: Cambridge University Press.

Maciel, C., & Pereira, V. C. (2013). *Digital legacy and interaction. Post-mortem issues.* Cham, Switzerland: Springer.

Magnani, L. (2013). Is abduction ignorance-preserving? Conventions, models and fictions in science. *Logic Journal of IGPL*, *21*(6), 882–914.

Marcus, A. (1992). *Graphic design for electronic documents and user interfaces.* New York, NY: ACM.

Mullet, K., & Sano, D. (1995). *Designing visual interfaces: communication oriented techniques.* Upper Saddle River, NJ: Sunsoft Press.

Nadin, M. (1988a). Interface design and evaluation—semiotic implications. In H. R. Hartson & D. Hix (Eds.), *Advances in human-computer interaction* (Vol. *2*, pp. 45–100). Norwood, NJ: Ablex.

Nadin, M. (1988b). Interface design: A semiotic paradigm. *Semiotica*, *69*(3–4), 269–302.

Nadin, M. (2001). One cannot not interact. *Knowledge-Based Systems*, *14*, 437–440.

Nadin, M. (2007). Semiotic machine. *Public Journal of Semiotics*, *1*(1), 57–75.

Nake, F. (2005). Computer art: a personal recollection. In *Proceedings of the 5th conference on Creativity and cognition* (pp. 54–62). New York, NY: ACM.

Nake, F., & Grabowski, S. (2001). Human-computer interaction viewed as pseudo-communication. *Knowledge-Based Systems*, *14*, 441–447.

Noble, J., & Biddle, R. (2002). Patterns as signs. In B. Magnusson (Ed.), *ECOOP 2002—Object-oriented programming* (pp. 368–391). Berlin: Springer.

Norman, D. A. (2010). *Living with complexity.* Cambridge, MA: MIT Press.

Nöth, W. (1995). *Handbook of semiotics.* Bloomington, IN: Indiana University Press.

Nöth, W. (2002). Semiotic Machines. *Cybernetics and human knowing*, *9*(1), 5–21.

Nöth, W. (2011). The semiotic foundations of pragmatics. In W. Bublitz, & N. R. Norrick (Eds.), *Foundations of pragmatics* (pp. 167–202). Boston, MA: Walter de Gruyter.

O'Halloran, L. K. (2015). Multimodal Digital Humanities. In P. P. Trifonas (Ed.), *International handbook of semiotics* (pp. 389–415). Dordrecht: Springer Netherlands.

O'Neill, S. (2008). *Interactive media the semiotics of embodied interaction*. London: Springer.

O'Neill, S., Benyon, D. R., & Turner, S. R. (2002). Semiotics and interaction analysis. In S. Bagnara, S. Pozzi, A. Rizzo, & P. Wright (Ed.), *Eleventh European Conference on Cognitive Ergonomics (ECCE 11): Cognition, culture and design* (pp. 44–50). Roma: Instituto di Scienze e Tecnologie della Cognizione / Conziglio Nazionale delle Ricerche.

Paternò, F., & Wulf, V. (2017). *New Perspectives in End-User Development*. Cham: Springer International Publishing AG.

Peirce, C. P. (1992). *The essential Peirce—selected philosophical writings Vol. I (1867–1893)* (N. Houser, & C. Kloesel, Eds.). Bloomington, IN: Indiana University Press.

Peirce, C. P. (1998). *The essential Peirce—selected philosophical writings Vol. II (1893–1913)* (N. Houser, & C. Kloesel, Eds.). Bloomington, IN: Indiana University Press.

Pereira, R., & Baranauskas, M. C. (2015). A value-oriented and culturally informed approach to the design of interactive systems. *International Journal of Human-Computer Studies*, *80*, 66–82.

Pereira, V. C., & Maciel, C. (2013). The internet generation and the posthumous interaction. In C. Maciel, & V. C. Pereira (Eds.), *Digital legacy and interaction* (pp. 63–81). Cham, Switzerland: Springer.

Prates, R. O., Rosson, M. B., & de Souza, C. S. (2015a). Interaction anticipation: Communicating impacts of groupware configuration settings to users. In *End-user development. Proceedings of Fifth International Symposium, IS-EUD 2015. 9083* (pp. 192–197). Cham, Switzerland: Springer International Publishing.

Prates, R. O., Rosson, M. B., & de Souza, C. S. (2015b). Making decisions about digital legacy with Google's inactive account manager. In J. Abascal, S. Barbosa, M. Fetter, T. Gross, P. Palanque, & M. Winckler (Eds.), *Human-computer interaction—INTERACT 2015. 9296* (pp. 201–209). Cham, Switzerland: Springer International Publishing.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. New York, NY: Cambridge University Press.

Salgado, L. C., Leitão, C. F., & de Souza, C. S. (2013). *A journey through cultures—metaphors for guiding the design of cross-cultural interactive systems*. London: Springer-Verlag.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. London: Cambridge University Press.

Searle, J. R. (1975). Indirect speech acts. In P. Cole, & J. L. Morgan (Eds.), *Syntax and semantics (3): Speech acts* (Vol. *3*, pp. 59–82). New York, NY: Academic Press.

Sedda, F. (2015). Semiotics of culture(s): Basic questions and concepts. In P. P. Trifonas (Ed.), *International Handbook of Semiotics* (pp. 675–696). Dordrecht: Springer Netherlands.

Stamper, R., Liu, K., Hafkamp, M., & Ades, Y. (2000). Understanding the roles of signs and norms in organisations. *Journal of Behaviour and Information Technology*, *19*(1), 15–27.

Sundar, S. S., & Nass, C. (2000). Source orientation in human-computer interaction programmer, networker, or independent social actor. *Communication Research*, *27*(6), 683–703.

Tanaka-Ishii, K. (2010). *Semiotics of programming*. Cambridge: Cambridge University Press.

Tanaka-Ishii, K. (2015). Semiotics of Computing: Filling the Gap Between Humanity and Mechanical Inhumanity. In P. P. Trifonas (Ed.), *International handbook of semiotics* (pp. 981–1002). Dordrecht: Springer Netherlands.

Trifonas, P. P. (2015). *International handbook of semiotics*. Dordrecht: Springer Netherlands.

Tuerk, A. (2013, Thursday, April 11, 2013). Plan your digital afterlife with Inactive Account Manager.

Turkle, S. (2005). *The second self: Computers and the human spirit* (20th anniversary edition). Cambridge, MA: MIT Press.

Valtolina, S., Barricelli, B. R., & Dittrich, Y. (2012). Participatory knowledge-management design: A semiotic approach. *Journal of Visual Languages and Computing*, *23*(2), 103–115.

Winograd, T., & Flores, F. (1986). *Understanding Computers and Cognition: A New Foundation for Design*. Reading, MA: Addison-Wesley.

3

Benefiting from ISO Standards

Nigel Bevan and Jonathan Earthy

Introduction

What is an international standard?

The period since the late 1980s has seen the development of a wide range of international standards related to HCI and usability. Many of these are not standards in the sense that the word is commonly understood. There are few absolutes in designing for usability: it depends on factors such as the context of use, design environment, resources constraints, and importance of usability. Consequently, many of the standards for HCI contain guidance and recommendations on good practice.

Most international standards for HCI are developed under the auspices of the International Organization for Standardization (ISO) and the International Electrotechnical Commission (IEC).[1] The current status of the ISO and IEC standards included in subsequent tables is explained in Table 3.1.

The initial work on HCI standards was undertaken by the ISO TC 159/SC4 "Ergonomics of human-system interaction" subcommittee (see Health and Safety Executive, 2003), and many of these standards contain general principles from which appropriate interfaces and procedures can be derived. This makes the standards authoritative statements of good professional practice, but makes it difficult to know whether an interface conforms to the standard. Reed et al. (1999) discuss approaches to conformance in these standards.

The ISO/IEC JTC1 "Information technology" committee has more recently established SC 35 "User Interfaces", which evolved out of early work on keyboard layout. This group has produced more specific standards for user interfaces.

Usability experts have worked with the ISO/IEC JTC1/SC7 "Software and systems engineering" subcommittee to integrate usability into software engineering and software quality standards. This has required some compromises: for example reconciling different definitions of usability by adopting the new term "quality in use" to represent the ergonomic concept of usability (Bevan, 1999).

[1] The phrase "ISO standards" is used in this chapter to refer generally to standards published by ISO, IEC, and ISO/IEC.

The Wiley Handbook of Human Computer Interaction, Volume 1, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

Table 3.1 ISO and IEC document types.

Example	Explanation
ISO 1234 2016	ISO standard 1234, published in 2016
ISO 1234-1 2016	Part 1 of ISO standard 1234
ISO/IEC 1234 2016	Joint ISO/IEC standard 1234
ISO 1234 2016 TS	An ISO Technical Specification: a normative document that may later be revised and published as a standard.
ISO 1234 2016 PAS	An ISO Publicly Available Specification: a normative document with less agreement than a TS, which may later be revised and published as a standard.
ISO 1234 2016 TR	An ISO Technical Report: an informative document containing information of a different kind from that normally published in a normative standard.
ISO 1234 XX	A draft standard XX — Document type DIS — Draft International Standard PDTS — Proposed Draft Technical Specification CD — Committee Draft AWI — Approved Work Item NP — New Proposal

Types of standard for HCI

Standards related to HCI and usability can be categorized as primarily concerned with:

- The user interface and interaction.
- Usability and human-centered quality (quality in use) as an intended outcome.
- The human-centered design process used to develop the product.
- The capability of an organization to apply human-centered design.

Figure 3.1 illustrates the logical relationship between the different types of standards. The intended outcome is usability: for the product to be effective, efficient, and satisfying when used in the intended contexts. A prerequisite for this is an appropriate interface and interaction for the product, system or service. This requires a human-centered design process, which to be achieved consistently requires an organizational capability to support human-centered design.

The User Interface and Interaction

Ergonomic guidelines

Most of the early effort in ergonomic standards for usability went into producing detailed guidelines for user interface design in the ISO 9241 series (Stewart, 2000), in the same tradition as the Smith and Mosier (1986) guidelines. The exhaustive ISO 9241 user interface guidelines (Table 3.2) and other standards produced by ISO TC159/SC4 (Table 3.3) include the design of presentation of information, user guidance, menus, direct manipulation, forms, haptic interfaces and accessibility.

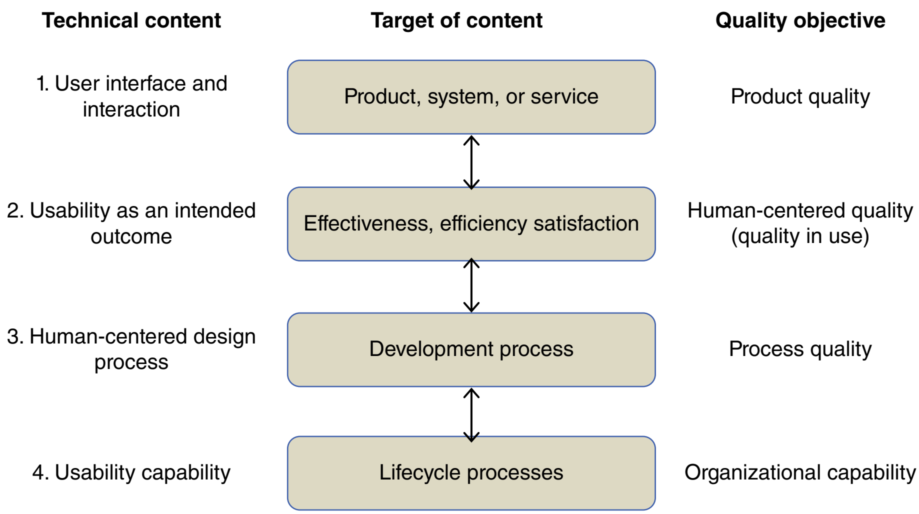


Figure 3.1 Categories of standard.

Table 3.2 ISO 9241 "Ergonomics of human-system interaction" standards.

Part	Date	ISO 9241 "Ergonomics of human-system interaction" standards
1	1997	General introduction
2	1992	Guidance on task requirements
5	1998	Workstation layout and postural requirements
6	1999	Guidance on the work environment
11	1998	Guidance on usability
11	2018	Usability: Definitions and concepts
13	1998	User guidance
14	1997	Menu dialogues
20	2008	Accessibility guidelines for ICT equipment and services
100	**2010**	**Introduction to standards related to software ergonomics (TR)**
110	2006	Dialogue principles
112	2017	Principles for the presentation of information
125	2017	Guidance on visual presentation of information
129	2010	Guidance on software individualization
143	2012	Forms
154	2013	Interactive voice response (IVR) applications
161	2016	Guidance on visual user interface elements
171	2008	Guidance on software accessibility
210	2010	Human-centered design for interactive systems
220	2017 DIS	Processes for enabling, executing and assessing human-centered design within organizations
300-392		**Requirements for electronic visual display screens**
400	**2007**	**Principles and requirements for physical input devices**
410	2008	Design criteria for physical input devices
411	2012	Evaluation methods for the design of physical input devices (TS)
420	2011	Selection of physical input devices
910	**2011**	**Framework for tactile and haptic interaction**
920	2009	Guidance on tactile and haptic interactions
940	2017	Evaluation of tactile and haptic interactions
960	2017	Framework and guidance for gesture interactions

Table 3.3 ISO TC159/SC4 "Ergonomics of human-system interaction" standards.

ISO number	ISO TC159/SC4 "Ergonomics of human-system interaction" standards
9241 (cf. Table 3.2)	Ergonomics of human-system interaction
9355 (3 parts)	Ergonomic requirements for the design of displays and control actuators
11064 (7 parts)	Ergonomic design of control centers
14915 (3 parts)	Software ergonomics for multimedia user interfaces
16982 2002 TR	Usability methods supporting human-centered design
18152 2010 TS	A specification for the process assessment of human-system issues
18529 2000 TS	Human-centered lifecycle process descriptions
20278 2015 TR	Unwanted reflections from the active and inactive areas of display surfaces visible during use
20282-1 2006	Ease of operation of everyday products—Part 1 Design requirements for context of use and user characteristics
20282-2 2013 TS	Usability of consumer products and products for public use—Part 2 Summative test method
21144 2016 TS	Ergonomics of human-system interaction — Electronic paper display—Indoor use

They provide comprehensive reference material, and can be used as an authoritative justification for design decisions. Some are structured as conditional guidelines (Harker, 1995).

But while the ISO 9241 user interface guidelines constitute an immense body of knowledge, they are not very easy for designers to use (Carter, 1999; de Souza and Bevan, 1990; Thovtrup & Nielsen, 1991). In the case of web design, the U.S. Department of Health and Human Services (HHS) had the resources to develop and maintain a set of guidelines (U.S. Department of Health and Human Services, 2016a) that were superior in presentation and content to the original ISO standard (ISO 9241-151) (Bevan and Spinhof, 2007), which has recently been withdrawn.

ISO 9241-110 "Dialogue principles" describes more general ergonomic design principles and provides a framework for applying those principles to the analysis, design, and evaluation of interactive systems. The intention of the principles is similar to Jacob Nielsen's 10 usability heuristics for user interface design (Nielsen, 1995).

ISO 9241 also provides requirements for physical input devices and display screens.

Interface elements

ISO/IEC JTC1/SC35 User Interfaces has developed more detailed standards specifying aspects of particular types of user interface (Table 3.4).

The types of standards developed by each SC 35 working group include:

WG 1 Keyboards, methods, and devices related to input and its feedback

- Layouts for different types of keyboards. New work is starting on guidance on virtual (e.g. onscreen) keyboards.
- A framework for gestures to support interoperability across various input devices and methods.

Table 3.4 ISO/IEC JTC1/SC35 standards (for accessibility see Table 3.5).

ISO/IEC number	ISO/IEC JTC 1/SC 35 "User interfaces" standards
WG1 Keyboards, methods and devices related to input and its feedback	
9995 (11 parts)	Keyboard layouts for text and office system
10741-1 1995	Cursor control for text editing
14754 1999	Pen-based interfaces—Common gestures for text editing with pen-based systems
14755 1997	Input methods to enter characters from the repertoire of 10646 with a keyboard or other input device
15411 1999	Segmented keyboard layouts
15412 1999	Portable computer keyboard layouts
15440 2016 TR	Future keyboards and other input devices and entry methods
24757 2008	Keyboard interaction model—Machine-readable keyboard description
30113 (3 parts)	Gesture-based interfaces across devices and methods
WG2 Graphical user interface and interaction	
11581	Icon symbols and functions (8 parts)
13251 2017 DIS	Collection of graphical symbols for office equipment
18035 2003	Icon symbols and functions for controlling multimedia software applications
18036 2003	Icon symbols and functions for World Wide Web browser toolbars
24738 2006	Icon symbols and functions for multimedia link attributes
24755 2007	Screen icons and symbols for personal mobile communication devices
WG4 User interfaces for mobile devices	
18021 2002	User interfaces for mobile tools for management of database communications in a client-server model
17549 (2 parts)	User interface guidelines on menu navigation
WG5 Cultural and linguistic adaptability	
15897 2011	User interfaces—Procedures for the registration of cultural elements
19764 2005 TR	Guidelines, methodology and reference criteria for cultural and linguistic adaptability in information technology products
20007 2014 TR	Cultural and linguistic interoperability—Definitions and relationship between symbols, icons, animated icons, pictograms, characters, and glyphs
20382 (2 parts)	User interface—Face-to-face speech translation
24785 2009 TR	Taxonomy of cultural and linguistic adaptability user requirements
30109 2015 TR	User interfaces—Worldwide availability of personalized computer environments
30112 2014 TR	Specification methods for cultural conventions
30122 (4 parts)	User interfaces—Voice commands
WG7 User interfaces object, actions and attributes	
11580 2007 TR	Framework for describing user interface objects, actions and attributes
WG8 User interfaces for remote interactions	
24752 (7 parts)	User interfaces—Universal remote console
Internet of Things (IoT) User interfaces	

WG 2 Graphical user interface and interaction

- Collections of symbols and icons.
- Guidelines for the design of accessible icons and symbols.

WG 4 User interfaces for mobile devices

- Guidelines on the design of navigation methods for using four-direction keys to make menu selections.
- User interface functions for management of mobile device communication.

WG 5 Cultural and linguistic adaptability

- A range of standards to support integration of the requirements for cultural and linguistic adaptability and user interfaces (CLAUI) aspects in products standards so that software and hardware can be adapted to local cultural and linguistic needs.

WG 6 User interfaces accessibility

- Several standards to support user interface accessibility, including: a usability code of practice, a collection of user accessibility needs, accessibly of various types of interface components, a common access profile and guidance on IT interoperability with assistive technology.

(ISO 9231-20 and ISO 9241-172 (Table 3.2) are also concerned with accessibility, and several other ISO TC159/SC4 standards (Table 3.3) are concerned with accessible design.

WG 7 User interfaces objects, actions and attributes

- Guidance both on the standardization of user interface objects, actions, and attributes, and on the implementation of these objects, actions, and attributes in any or all modalities.

WG 8 User interfaces for remote interactions

- A standard for "universal remote consoles" that supports the development, deployment and use of pluggable and personal user interfaces that can provide "Accessible User Interfaces, for Everyone, Everywhere, on Anything" (see Bund et al., 2010; Klima et al. 2009).

AHG 1 Internet of Things (IoT) user interfaces New work is starting in this area.

Accessibility

Accessibility standards have been developed by both ISO and ISO/IEC JTC 1 (Table 3.5).

Other applications of ergonomics

ISO TC159 and other ISO committees have developed standards for specific applications of ergonomics (Table 3.6).

Table 3.5 Accessibility standards.

ISO number	
ISO/IEC JTC1/SC35	
13066 (5 parts)	Interoperability with assistive technology (AT)
19765 2007 TR	Survey of icons and symbols that provide access to functions and facilities to improve the use of information technology products by the elderly and persons with disabilities
19766 2007 TR	Guidelines for the design of icons and symbols accessible to all users, including the elderly and persons with disabilities
20071 (4 parts) TS	User interface component accessibility
24756 2009	Framework for specifying a common access profile (CAP) of needs and capabilities of users, systems, and their environments
24786 2009	Accessible user interface for accessibility settings
29136 2012	Accessibility of personal computer hardware
29138 (3 parts) TR	Accessibility considerations for people with disabilities
30071-1 NP	Development of user interface accessibility—Part 1: Code of practice for creating accessible ICT products and services (including accessible Web products)
ISO TC159	
9241-20 2008	Accessibility guidelines for ICT equipment and services
9241-171 2008	Guidance on software accessibility
21054-1 AWI TS	Input controls of consumer products—Part 1 Input controls accessibility for basic operation
21055 2017 CD	Minimum legible font size for people at any age
21056 2017 DIS	Guidelines for designing tactile symbols and letters
24500 2010	Accessible design—Auditory signals for consumer products
24501 2010	Sound pressure levels of auditory signals for consumer products
24502 2010	Specification of age-related luminance contrast for colored light
24503 2011	Tactile dots and bars on consumer products
24504 2014	Sound pressure levels of spoken announcements for products and public address systems
24505 2016	Method for creating color combinations taking account of age-related changes in human color vision
24507 2017 DIS	Doors and handles of consumer products
Other ISO	
11608-7 2016	Needle-based injection systems for medical use—Requirements and test methods—Part 7: Accessibility for persons with visual impairment
21452 2011	Building construction—Accessibility and usability of the built environment
Other ISO/IEC JTC1	
Guide 71 2014	Guide for addressing accessibility in standards
10779 2008	Office equipment accessibility guidelines for elderly persons and persons with disabilities
13066 (4 parts) TR	Interoperability with assistive technology (AT)
20016-1 2014	Language accessibility and human interface equivalencies (HIEs) in e-learning applications
24751 (3 parts) 2008	Individualized adaptability and accessibility in e-learning, education, and training
40500 2012	W3C Web Content Accessibility Guidelines (WCAG) 2.0

Table 3.6 Standards for other applications of ergonomics.

ISO or ISO/IEC number	
Road vehicles (see Chapter 40 of this handbook: Motor Vehicle Driver Interfaces)	
(12 standards)	Ergonomic aspects of transport information and control systems
(4 standards)	Ergonomic requirements for the driver's workplace in line-service buses
(4 standards)	Ergonomic aspects of in-vehicle presentation for transport information and control systems
16505 2015	Ergonomic and performance aspects of camera monitor systems
Danger signals	
7731 2003	Danger signals for public and work areas—Auditory danger signals
11428 1996	Visual danger signals—General requirements, design, and testing
11429 1996	System of auditory and visual danger and information signals
Other standards	
6385 2016	Ergonomics principles in the design of work systems
9886 2004	Evaluation of thermal strain by physiological measurements
9921 2003	Assessment of speech communication
10075 (3 parts)	Ergonomic principles related to mental workload
15534 (3 parts)	Ergonomic design for the safety of machinery
16976 (8 parts)	Respiratory protective devices—Human factors
19358 2002	Construction and application of tests for speech technology
24500-5 2010	Accessible design—Auditory signals for consumer products
(5 standards)	Ergonomics of the thermal environment

Usability and Human-centered Quality as Intended Outcomes

Usability and human-centered quality

While the everyday meaning of usability is ease of use, in 1998 ISO 9241-11 gave usability a more strategic interpretation as the intended outcome of interaction: the extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use. This transforms usability from something nice to have to a strategic business objective: it is important that users can complete tasks successfully in an acceptable amount of time, and in the Internet age it is essential that customers are satisfied.

This approach to usability can encompass accessibility (effectiveness, efficiency, and satisfaction for people with the widest range of capabilities) and user experience (UX). User experience has many interpretations (see Chapter 10 of this handbook), but is defined in ISO 9241-210 as a person's perceptions and responses resulting from the use and/or anticipated use of a product, system or service (including all the users' emotions, beliefs, preferences, perceptions, physical and psychological responses, behaviors, and accomplishments that occur before, during, and after use). The current revision of ISO 9241-11 extends usability to include achieving both pragmatic and hedonic goals, thus including many aspects of user experience.

ISO 9241-220 has recently introduced the new term "human-centered quality" to combine these properties together with the role of usability in minimizing any negative consequences and optimizing positive outcomes. This recognizes that it is not sufficient simply to consider successful interaction—it is also important to minimize the possibility of any adverse consequences that could arise from poor usability. (Examples could include an individual failing to purchase the correct transport ticket, adverse economic consequences for a company resulting from user errors, and environmental consequences resulting from poor usability of home-heating controls.) ISO 9241-220 explains that human-centered quality is the major objective of human-centered design. Communicating usability in terms of the benefits provided by the components of human-centered quality can provide convincing justification for the resources needed for human-centered design activities (see Figure 3.2, adapted from ISO 9241-220).

Usability and software quality

In the original ISO/IEC 9126:1990 software quality standard, usability referred only to the capability of the user interface to be usable. However, the broad ergonomic view of usability was subsequently incorporated into the software quality standards as "quality in use" (Bevan, 1999). This has the major benefit of making the ergonomic concept of usability (reconceptualized as the quality experienced by the user) an objective for software engineering.

ISO/IEC 25010:2011 (Table 3.7), which has replaced ISO/IEC 9126, defines quality in use as: "the degree to which a product or system can be used by specific users to meet their needs to achieve specific goals with effectiveness, efficiency,

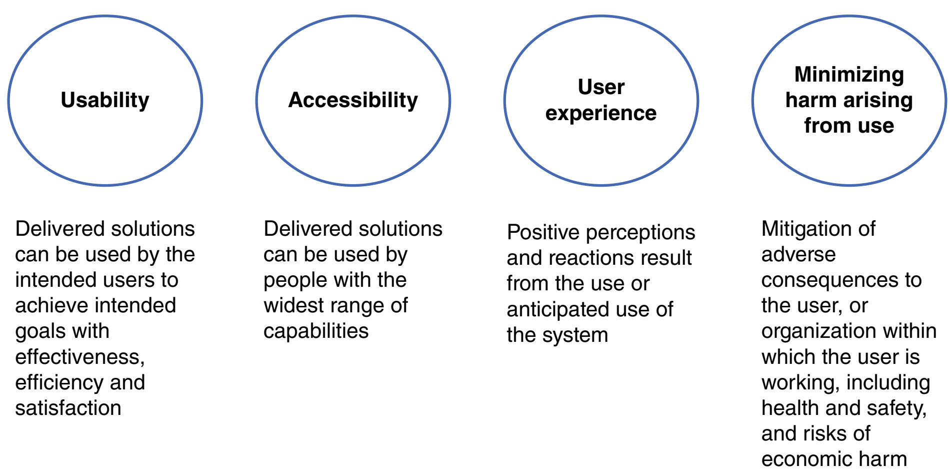


Figure 3.2 Human-centered quality.

Table 3.7 ISO/IEC 250xx Quality model and measures standards related to usability.

ISO/IEC number	Systems and Software Quality Requirements and Evaluation (SQuaRE)
25010 2011	System and software quality models
25011 2017 TS	Service quality model
25022 2016	Measurement of quality in use
24023 2016	Measurement of system and software product quality

freedom from risk and satisfaction in specific contexts of use," which is very similar to the new concept of human-centered quality. Discussions are taking place to see whether a closer alignment of these concepts is possible.

Assuring usability

The intention when ISO 9241-11 was first drafted in 1988 was to produce "usability assurance statements" consisting of test reports giving results for effectiveness, efficiency, and satisfaction in a specified context of use (Bevan & Holdaway, 1993). Unfortunately some parts of industry at that time did not want to be forced to produce usable products, and a large U.S. company threatened to use its influence to ensure that the standard was rejected unless it was redrafted as guidance rather than requirements.

However, the same concept was reinvented in the United States 10 years later as the Common Industry Format for Usability Test Reports by a group of companies frustrated by the low profile of usability in product procurement (Bevan et al., 2002). This became the US standard ANSI/NCITS 354 in 2001, and the international standard ISO/IEC 25062 in 2006. ISO/IEC 25062 specifies the contents to be included in a report of summative usability testing that would enable a supplier to demonstrate the usability of their system to a potential purchaser, and the purchaser to judge whether the system would be usable in their own context of use.

ISO 20282-2 specifies a test method for summative usability testing that can produce results that are sufficiently reliable to provide usability assurance for consumer products and products for public use. ISO 20282 was published as a preliminary standard for review in several parts in 2006 and 2007, but not before a large German company had lobbied hard to prevent publication of a standard that might be used to regulate the usability of consumer products.

Usability of medical devices

The original IEC 62366:2007 standard has recently been withdrawn, and replaced by two new parts:

- IEC 62366-1:2015 "Application of usability engineering to medical devices" specifies a process for a manufacturer to analyze, specify, develop, and evaluate the usability of a medical device as it relates to safety. This usability engineering process enables the manufacturer to assess and mitigate risks associated with correct use and use errors. It can also be used to identify risks associated with abnormal use.
- IEC TR 62366-2:2016 "Guidance on the application of usability engineering to medical devices" provides guidance that addresses specific areas that can be helpful for those implementing a usability engineering process both as defined in IEC 62366-1:2015 and for supporting goals other than safety.

Measures of usability and quality in use

ISO/IEC 25022 (Table 3.7) produced by ISO/IEC JTC1/SC7 Software Engineering specifies measures of quality in use that include measures for the same components of usability that are defined in ISO 9241-11: effectiveness, efficiency, and satisfaction (Table 3.8).

Table 3.8 Measures of effectiveness, efficiency, and satisfaction.

Effectiveness	Efficiency	Satisfaction
Tasks completed	Task time	Overall satisfaction
Objectives achieved	Time efficiency	Satisfaction with features
Errors in a task	Cost effectiveness	Discretionary usage
Tasks with errors	Productive time ratio	Feature utilization
Task error intensity	Unnecessary actions	Proportion of users complaining
	Fatigue	Proportion of user complaints about a particular feature
		User trust
		User pleasure
		Physical comfort

Table 3.9 Measures of risk.

Economic risk	Health and safety risk	Environmental risk
Return on investment (ROI)	User health reporting frequency	Environmental impact
Time to achieve return on investment	User health and safety impact	
Business performance	Safety of people affected by use of the system	
Benefits of IT Investment		
Service to customers		
Website visitors converted to customers		
Revenue from each customer		
Errors with economic consequences		

ISO/IEC 25022 also suggests measures for freedom from risk (Table 3.9), which it defines as the "degree to which the quality of a product or system mitigates or avoids potential risk to the user, organization or project, including risks to economic status, human life, health, or the environment." This is similar to "minimizing any negative consequences" in human-centered quality.

An example measure is: "Proportion of usage situations where there are human or system errors with economic consequences." Although it is usually not possible to control other factors in the context of use that could influence freedom from risk, it is often possible to provide evidence for the potential risks that could result from poor usability or poor product quality.

ISO/IEC 25022 also provides measures for context coverage: "the degree to which a product or system can be used with effectiveness, efficiency, satisfaction, and freedom from risk in both specified contexts of use and in contexts beyond those initially explicitly identified" (Table 3.10). Context coverage was introduced into ISO/IEC 25010 to provide for specification and evaluation of usability in all defined and anticipated contexts of use.

Flexibility measures are used to assess the degree to which a product or system can be used with acceptable levels of effectiveness, efficiency, satisfaction, and freedom from risk in contexts beyond those initially specified in the requirements for the system. Flexibility enables products to take account of circumstances, opportunities, and individual preferences that might not have been anticipated in advance.

Table 3.10 Measures of context coverage.

Context completeness	Flexibility
Context completeness	Flexible context of use Product flexibility Proficiency independence

Proficiency independence assesses the extent to which the product can be used by people who do not have specific knowledge, skills, or experience.

Human-Centered Design

Human-centered design process

The overview of the human-centered design processes for interactive systems was originally published as ISO 13407 in 1999, and in 2010 the content was revised and renumbered as ISO 9241-210. Intended as a managers' guide, it is one of the best concise introductions to usability and human-centered design that is available. ("Human" centered rather than "user" centered to acknowledge the importance of stakeholders who may not be users.)

Some of the recommendations in ISO 13407 have been made requirements. For example, for a development process to show conformance with ISO 9241-210, the requirements include:

- Project planning shall allocate time and resources for the human-centered activities. This shall include time for iteration and the incorporation of user feedback, and for evaluating whether the design solution satisfies the user requirements.
- Relevant user and stakeholder groups shall be identified and their relationship with the proposed development described in terms of key goals and constraints.
- There are four linked human-centered design activities that shall take place during the design of any interactive system.
  - understand and specify the context of use;
  - specify the user requirements;
  - produce design solutions;
  - evaluate.

This makes ISO 9241-210 a powerful tool to assure the basic elements of a human-centered design process.

Common Industry Format for documenting usability

The ISO/IEC 2506x Common Industry Format standards for documenting usability as part of human-centered design (Table 3.11) are being developed jointly by the ISO TC159 Ergonomics and JTC1 Information Technology committees.

Table 3.11 ISO/IEC 2506x "Common Industry Format (CIF) for usability" standards.

ISO/IEC number		Common Industry Format (CIF) for usability standards
25060	2010	General framework for usability-related information
25062	2006	Usability test reports
25063	2014	Context of use description
24064	2013	User needs report
25065	CD	User requirements specification
25066	2016	Evaluation report
2506n	planned	User interaction specification
2506m	planned	User interface specification

ISO/IEC 25062: Usability test reports Defines the content of summative usability evaluation reports.

ISO/IEC 25063: Context of use description Defines the information about the context of use (the characteristics of the users, tasks, and environments) that should be identified and documented at different stages of design and development.

ISO/IEC 25064: User needs report Defines the types of information that should be included in a user needs report, including:

- user responsibilities and goals;
- source data on which user needs are based;
- identified and reported user needs, expressed in terms of the user or the set of users that it relates to, the intended outcome to be achieved, the prerequisite (need) identified as necessary to achieve the intended outcome, and the specific context of use in which it applies;
- identified and reported management and other stakeholder needs;
- performance deficiencies/problems/potential improvements (if identified).

ISO/IEC CD 25065: User requirements specification Defines the information that should be included in a user requirements specification, including the human-centered quality objectives, detailed user requirements expressed in terms of the things that a user shall be able to do, the intended context of use and any known constraints.

ISO/IEC 25066: Evaluation report

- Specifies what should be included in usability evaluation reports. It identifies different types of usability evaluation and specifies the contents that should be included in reports for each type of usability evaluation:
  - Inspection to identify usability defects and the corresponding potential usability problems.
  - User observation
    - Qualitative: Observing user behavior to identify actual usability problems.
    - Quantitative: Measuring user performance and responses to obtain data on effectiveness and efficiency.
  - Obtaining subjective information from users including:
    - Qualitative: Problems, opinions, and impressions given during or after a usability evaluation.
    - Quantitative: Measures of user satisfaction or perception.

The Capability of an Organization to Apply User-centered Design

Human-centered design process models

Two comprehensive models of human-centered design were developed based on ISO 13407, and these still represent the most comprehensive published models of good practice (Earthy, Sherwood Jones, & Bevan, 2001). ISO TR 18529 (Table 3.3) covers broadly the scope of ISO 13407, structured as a set of 44 activities in seven categories. It was derived from surveys of good practice in industry, and has been used as the basis for assessing whether a project has adequately implemented human-centered design, for assessing the usability maturity of an organization (Bevan, 2005b; Earthy et al., 2001), and it provided the basis for a proposed scheme for accrediting usability professionals (Bevan, 2003).

ISO TS 18152, originally developed in conjunction with the UK defense industry, is wider in scope, covering the whole range of human-centered activities involved in systems engineering. It can be used in conjunction with the ISO/IEC 15288 systems engineering standard, which already has pointers to the essential human-centered activities.

ISO TS 18152 formed the basis for the structure of IEC 62508-2011 “Guidance on human aspects of dependability”, which provides an elaboration of the HCD activities for application to safety-related systems. IEC 62508 categorizes the ISO TS 18152 activities by project stages, and gives guidance on how they can be applied to achieve whole system integrity.

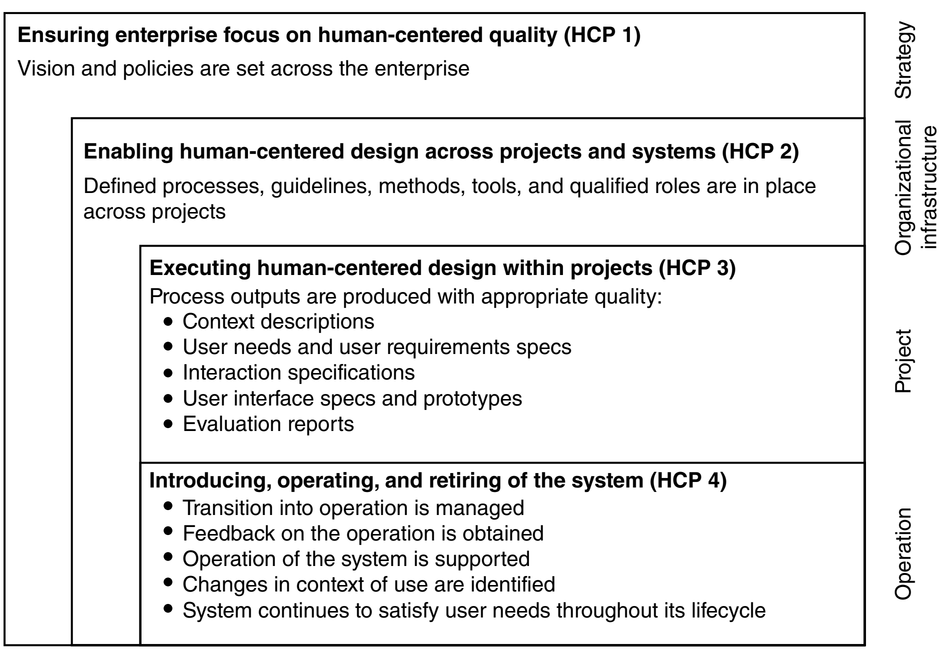


Figure 3.3 Human-centered design process categories and contents.

ISO TR 18529 has been replaced by ISO 9241-220 (Table 3.2). ISO 9241-220 elaborates on ISO 9241-210 to provide a comprehensive description of the processes that support the activities that are required as part of human-centered design. Figure 3.3 (adapted from ISO 9241-220) summarizes the processes that need to be in place in each area of an organization that has some responsibility for human-centered design. The groups of processes related to these levels are called “Human-Centered Process categories” (HCP). Together the implementation of these four sets of processes ensures that the systems produced, acquired, and operated by an organization have appropriate levels of usability, accessibility, user experience, and mitigation of risks that could arise from use.

ISO 9241-220 can be used for:

- implementing human-centered design based on the process outcomes needed to achieve human centered quality as part of a system development or procurement process, and / or support lifecycle;
- assessing an enterprise’s existing capability to carry out the human-centered processes;
- improving the effectiveness of human-centered design as part of an existing system development process;
- specification and development of necessary competence in human-centered design.

Assessing usability capability

Usability maturity assessment can be used to profile the capability of an organization to take account of human-centered issues in all relevant design, development, and support activities. By identifying the strengths and weaknesses of an organization it is possible to identify potential areas for improvement, and to suggest the most cost-effective methods and techniques that could be used to improve the capability.

Each process shown in Figure 3.3 is composed of a set of outcomes and activities needed to achieve the objectives of the processes. The 148 outcomes identified in the model can be used as a checklist of good practice in human-centered design. The model is tailored to the needs of an organization before use, eliminating any processes that are either not relevant to the business, or outside the scope of the assessment.

The ISO/IEC 33000 series of standards provide methods for carrying out process assessment. The results of assessing the extent to which each process is carried out can be used to identify cost-effective methods and techniques that can be integrated with the existing system development processes to improve the usability capability.

How can International Standards for HCI be Used?

Standards have most impact when they are called up in legislation or a contract. Although some ergonomics standards for hardware can be used in support of health and safety regulations in the EU (Bevan, 1991; Health and Safety Executive, 2003), usability standards are likely to have most influence when they are cited in commercial contracts.

- The ultimate goal is to achieve usability when in use (i.e. effectiveness, efficiency, and satisfaction). ISO/IEC 25062 can be used to establish requirements for usability (Bevan et al., 2002) and to document whether the requirements have been met in a usability test.
- An organization could be required to demonstrate its usability capability, based on ISO 18152 or ISO 9241-220.
- A design and development project could be required to carry out activities that conform to ISO 9241-210.
- Interface design could be required to comply with the user interface guidelines in ISO 9241 parts 13 to 171.
- User interfaces can be required to conform to the specific requirements of the TC 35 standards.

All these standards could be used as a basis for education and training, and ISO 18152 or the new ISO 9241-220 can provide a framework for usability process improvement (Bevan, 2005b).

Development of ISO Standards

ISO and IEC comprise national standards bodies from member states. International standards are produced by groups of experts after a rigorous review process, and represent a consensus on the current state of the art. The technical work takes place in working groups of experts, nominated by national standards committees. Because of the number of international experts involved in their development they provide a more balanced perspective than is typically found in textbooks or individual publications.

The standards are developed over a period of several years, and in the early stages the published documents may change significantly from version to version until consensus is reached. As the standard becomes more mature, from the committee draft stage onwards, formal voting accompanied by comments takes place by participating national member bodies.

As ISO standards proceed through the draft stages (Table 3.1), they are circulated to participating national bodies and liaison organizations for comment. One way to contribute is by making comments through your national standards body or through an organization such as UXPA that has a liaison to ISO TC159/SC4. The working group is obliged to consider every submitted comment and to provide a disposition.

ISO standards are written by experts nominated by ISO national standards bodies and liaison organizations. The experts who write standards give their time free of charge, although some ISO national member bodies contribute to experts' travel expenses. A minimum of five countries must nominate experts for development of a new standard to be approved, and there is always a need for more suitably qualified experts who are available to contribute. So most national standards bodies welcome participation by additional experts.

Obtaining ISO Standards

The introduction, scope, and definitions in ISO standards can be previewed on the ISO Web site (www.iso.org). Some countries translate ISO standards into their national language. Standards can be purchased from ISO or national standards bodies as pdf files or on paper. For national and international standards bodies, sale of standards is an important source of income. Unfortunately, this makes them expensive to purchase individually (for example the 32 page ISO 9241-210 costs about U.S. $140). But many commercial organizations and educational institutions maintain a library of standards for use by their staff and students. International standards are unique in that they represent a consensus view that cannot be obtained elsewhere.

Conclusions

The majority of early effort in ergonomics standards went into developing conditional guidelines. The original parts 12 to 17 of ISO 9241 contained an overwhelming 82 pages of guidelines. At the time, several checklists were prepared to help assess conformance of software to the main principles in ISO 9241 (Gediga, Hamborg, & Düntsch, 1999; Oppermann & Reiterer, 1997; Prümper 1999). The HHS has faced the same problem with its Web guidelines. The latest version has a daunting 195 items. Probably for this reason, 41 items have been prioritized as standards that are required for the design and development of all US HHS/OS and priority Web sites (U.S. Department of Health and Human Services, 2016b).

While user interface standards remain important, there has been increasing emphasis on the value of using an appropriate human-centered design process to achieve usability and human-centered quality. Iterative design based on an in-depth understanding of users, their needs and goals, and the context in which the system will be used is likely to make a more substantial contribution to usability than the specific details of the user interface elements. ISO 9241-11, ISO 9241-210 and ISO 9241-220 provide an important foundation for both the process to be followed and the capability needed to implement the process.

Standards are more widely accepted in Europe than in the United States, partly for cultural reasons, and partly to achieve harmonization across EU countries. Many international standards (including ISO 9241) have been adopted as European standards. The EU Supplier's Directive (Council Directive 2004/18/EC on the coordination of procedures for the award of public works contracts, public supply contracts and public service contracts [2004] OJL 134, 30/04/2004, 114–120) requires that the technical specifications used for public procurement must be in terms of any relevant European standards. Ergonomic standards such as ISO 9241 can also be used to support adherence to European regulations for the health and safety of display screens (Bevan, 1991; Council Directive 90/270/EEC on the minimum safety and health requirements for work with display screen equipment [1990] OJ L 156, 21/06/1990, 14–18; Stewart, 2000).

However, it is not clear how many of the standards listed in this chapter are widely used. One weakness of some of the HCI standards is that up to the voting stage they have been developed collectively by experts without any user testing of prototypes during development. Exceptions include the US Common Industry Format that underwent trials during its evolution outside ISO, and ISO TS 18152 and ISO TR 18529, which were developed in a user-centered manner using UK and EU funding. There are ISO procedures to support incorporation of feedback, and ISO 20282 has been issued as a technical specification to encourage feedback before it is confirmed as a standard. This is an approach that should be encouraged in future.

Another potential weakness of international standards is that the development process can be slow, because the content depends on the voluntary effort of appropriate experts. For example, the original version of ISO 9241-11 took 10 years from conceptualization to publication. But ISO has increasingly put emphasis on speedy development, with most standards now being developed within 3 years, or a maximum of 4 years.

Acknowledgments

This chapter includes some updated material adapted from Bevan (2005a), Bevan (2009), and Bevan, Carter, Earthy, Geis, & Harker (2016). We thank Thomas Geis and Jim Carter for their feedback on the chapter.

References

ANSI (2001). Common industry format for usability test reports (ANSI-NCITS 354-2001). Washington, DC: American National Standards Institute.

Bevan, N. (1991). Standards relevant to European Directives for display terminals. In H. J. Bullinger (Ed.), *Proceedings of the 4th International Conference on Human Computer Interaction, Stuttgart*, September 1991. Amsterdam: Elsevier.

Bevan, N. (1999). Quality in use: Meeting user needs for quality. *Journal of Systems and Software*, *49*(1), f35.

Bevan, N. (2003). Accreditation of usability professionals. In J. Jacko, & C Stephanidis (Eds.), *Human-Computer Interaction: Theory and Practice (Part 1), Volume 1 of the Proceedings of HCI International* (pp. 429–433). Mawah, NJ: Lawrence Erlbaum.

Bevan, N. (2005a). International standards for HCI. *Encyclopedia of human computer interaction*. Hershey, PA: Idea Group Publishing. Retrieved from www.nigelbevan.com/papers/International_standards_HCI.pdf

Bevan, N. (2005b). Cost benefits framework and case studies. In R. G. Bias, & D. J. Mayhew (Eds.), *Cost-justifying usability: An update for the Internet age* (pp. 575–600). Burlington, MA: Morgan Kaufmann.

Bevan, N. (2009). International standards for usability should be more widely used. *Journal of Usability Studies*, *4*(3),106–113.

Bevan, N., Carter, J., Earthy, J., Geis, T., & Harker, S. (2016). New ISO standards for usability, usability reports and usability measures. In M. Kurosu (Ed.), *Proceedings, Part I, of the 18th International Conference on Human-Computer Interaction. Theory, Design, Development and Practice* (pp. 268–278). New York, NY: Springer-Verlag.

Bevan, N., Claridge, N., Maguire, M., & Athousaki, M. (2002). Specifying and evaluating usability requirements using the Common Industry Format: Four case studies. In *Proceedings of the IFIP 17th World Computer Congress, Montreal, Canada, 25–30 August 2002* (pp. 133–148). Amsterdam: Kluwer Academic Publishers.

Bevan, N., & Holdaway, K. (1993). User needs in information technology standards. In C. D. Evans, B. L. Meek & R. S. Walker (Eds.), *Technology standards*. London: Butterworth Heinemann.

Bevan, N., & Spinhof, L. (2007). Are guidelines and standards for web usability comprehensive? In J. Jacko (Ed.), *Human-Computer Interaction, Part I* (HCII 2007) (pp. 407–419). Berlin: Springer-Verlag.

Bund, J., Vanderheiden, G., Zimmermann, G., Zinnikus, I., Alexandersson, J., Klíma, M., … Urdaneta, E. (2010). OpenURC: standardization toward "user interfaces for everyone, everywhere, on anything." In *Access for All in the Desktop, Web and Mobile Field: An End-user and Developer Perspective. Proceedings of First International AEGIS Conference*, 7–8 October (pp. 268–275). Seville: University of Seville, Spain.

Carter, J. (1999). Incorporating standards and guidelines in an approach that balances usability concerns for developers and end users. *Interacting with Computers*, *12*, 179–206.

de Souza, F., & Bevan, N. (1990). The use of guidelines in menu interface design: Evaluation of a draft standard. In *Proceedings of IFIP INTERACT'90: Human-Computer Interaction 1990* (pp. 435–440). Amsterdam: North-Holland.

Earthy, J., Sherwood Jones, B., & Bevan, N. (2001). The improvement of human-centred processes—facing the challenge and reaping the benefit of ISO 13407. *International Journal of Human-Computer Studies*, *55*(4), 553–585.

Gediga, G., Hamborg, K., & Düntsch, I. (1999). The IsoMetrics usability inventory: An operationalisation of ISO 9241-10. *Behaviour and Information Technology*, *18*(3), 151–164.

Harker, S. (1995). The development of ergonomics standards for software. *Applied Ergonomics*, *26*(4), 275–279.

Health and Safety Executive (2003) Work with display screen equipment: Guidance on Regulations. Retrieved from http://www.hse.gov.uk/pUbns/priced/l26.pdf

Klíma, M., Macik, M., Urdaneta, E., Alexandersson, J., Buiza, C., Carrasco, E., … Alexandersson, J. (2009). In *Adjunct Proceedings of the Ambient Intelligence European Conference (AmI 2009)*, Salzburg, Austria (pp. 85–88). Salzburg: ICT&S Center, University of Salzburg.

Nielsen, J. (1995). *Ten usability heuristics for user interface design*. Retrieved from https://www.nngroup.com/articles/ten-usability-heuristics/

Oppermann, R., & Reiterer, R. (1997). Software evaluation using the 9241 evaluator. *Behaviour and Information Technology*, *16*(4/5), 232–245.

Prümper, P. (1999). Test it: ISONORM 9241/10. In Bullinger, H.-J. & Ziegler, J. (Eds.), In *Proceedings of HCI International, Munich, 22–27 August 1999*. Mahwah, NJ: Lawrence Erlbaum.

Reed, P., Holdaway, K., Isensee, S., Buie, E., Fox, J., Williams, J., Lund, A. (1999). *User Interface Guidelines and Standards: Progress, Issues, and Prospects. Interacting with Computers*, *12*(2), 119–142.

Smith, S. L., & Mosier, J. N. (1986). *Guidelines for designing user interface software*. (Report no. ESD-TR-86-278). Bedford, MA: The Mitre Corporation.

Stewart, T. (2000). Ergonomics user interface standards: are they more trouble than they are worth? *Ergonomics*, *43*(7), 1030–1044.

Thovtrup, H., & Nielsen, J. (1991). Assessing the usability of a user interface standard. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 335–341). New Orleans, LA: ACM.

U.S. Department of Health and Human Sciences (2016) Research-based web design and usability guidelines. Retrieved from http://webstandards.hhs.gov/guidelines/

U.S. Department of Health and Human Sciences (2016b) *Web standards*. Retrieved from https://webstandards.hhs.gov/standards/

4

Gender and Human-Computer Interaction

Samantha Breslin and Bimlesh Wadhwa

Introduction

What role does gender play in human-computer interaction (HCI)? In order to explore possible answers to this question it is also necessary to ask what is gender? And how does gender relate to technologies and computers? Research in fields such as anthropology, gender studies, feminist technology studies, and gender HCI has explored different ways gender is expressed and experienced by people around the world, and the intricate and intimate ways gender relates to technology design, development, and use, which are key to answering these questions. After discussing several problematic ways gender currently figures (largely implicitly) in HCI, this chapter provides an overview of theories and approaches to gender and technology, and steps towards designing gender-aware and gender-inclusive technologies. Gender HCI is a broad and multifaceted topic. It is an open question as to which approaches work best in particular circumstances. This chapter aims primarily to introduce multiple conceptual and practical tools for doing gender-sensitive HCI.

Gender sensitivity is significant for making technologies that facilitate usage and interaction for all genders. Yet, it has much broader implications—to encourage designers and developers to explore how values and norms affect technology design and development. Human-computer interaction is often conducted as an empirical scientific practice, using statistical analyses of usability and scientific models for cognition and psychological behavior, for example. Such approaches undoubtedly offer insights on how to enrich humans' interactions with computers. Yet, such interactions are also situated within historical and cultural contexts, which influence what technologies are designed, how, and for whom.

Problematic Gendering in HCI

There are several interrelated and problematic ways gender is implicated in HCI research and practice. Several of these problematics are at work even when gender is not being explicitly addressed as part of design or development. Gender is an intricate

The Wiley Handbook of Human Computer Interaction, Volume 1, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

part of everyday practice, affecting designers' interactions with one another, as well as the designs they create (see also Bath, 2014; Breslin & Wadhwa, 2014b).

Disparate numbers

Research since the 1980s has explored the dearth of women in computer science and related fields such as science, engineering, and mathematics, particularly in "Western" countries (Margolis & Fisher, 2002; Turkle, 1988). Recently, employee numbers released by large tech companies based in Silicon Valley show a similar lack of women and ethnic minorities, with 80–90% of tech positions filled by men and 92–94% by Whites and Asians at Google, Facebook, and Twitter, for example (https://diversity.google/commitments/; Van Huysse, 2014; Williams, 2014). This disparity cannot be explained by biological differences in relation to ability (Ceci, Williams, & Barnett, 2009). Yet, such disparities are highly problematic because they are often associated with a symbolic coding of technology and technological careers as masculine (Faulkner, 2001). This cyclically works to reinforce stereotypes about women's abilities. For example, studies have shown how women's performance often declines as a result of perceptions of being judged according to this stereotype (Spencer, Steele, & Quinn, 1999), as well as being judged for using technologies or programs seen as masculine (Cooper & Kugler, 2009). The disparity and symbolic coding can also lead to differentials in terms of access, support, and opportunity for women in using and designing technologies.

The I-methodology

In relation to HCI more specifically, this lack of diversity is also problematic due to a common practice by designers and developers of using themselves as model users, referred to as the "I-methodology" (Oudshoorn, Rommes, & Stienstra, 2004). Given that most developers, at least in the United States and United Kingdom, are educated, middle-class, White males, as just discussed, there is a tendency to reproduce the norms and values of that group (Berg, 1999; Forsythe, 2001; Oudshoorn et al., 2004). For example, the design of a smart house ignored much of the household work, which women tend to do, and was largely made for technically savvy male users, despite being intended as a product for "everyone" (Berg, 1999). There are many examples where women and others are excluded from designs and hindered in their abilities to use technologies because they are implicitly based on a model of a male or masculine user, such as voice-recognition software that does not recognize women's voices (Churchill, 2010) or Apple Health apps that allow users to track almost all aspects of their bodily functions, except for women to track their menstrual cycles (Quinn, 2014). This practice is likely one of the reasons recent research suggests that diverse teams are more economically successful, with the idea that diverse teams bring diverse ideas in designing and building technologies (Barker, Mancha, & Ashcraft, 2014).

Design stereotypes

Gendering through the I-methodology often occurs when gender is not explicitly considered as part of design. On the other hand, when gender is addressed explicitly, stereotyped ideas of gender are commonly used, or gender-sensitive design is equated with designing for women. These practices often result in products for women being simply made smaller or pink, or reinforcing gender norms with products based on fashion or sociality (Cassell, 2003; Churchill, 2010). This can work to “ghettoize” girls and women as a group that needs different technologies from “normal” ones, perpetuating stereotypes that girls or women do not like or are not good at technology (Cassell, 2003). Designing specifically for women, particularly based on stereotypes, does not help improve the gender inclusivity of shared technologies, and also excludes men as users from these products. It also ignores the significant diversity among women, including the differences between technically savvy women designers/developers and women users, as well as differences among women in relation to culture, race, sexuality, along with other categories of difference.

Gender difference

While recognizing differences is important to avoid homogenizing people, there has also been a tendency to rely on gender difference as a way to explain differences in numbers and usage in relation to technologies (Rode, 2011). For example, HCI research often parameterizes usability tests by gender in order to understand differences between men’s and women’s use (Beckwith & Burnett, 2004; Tan, Czerwinski, & Robertson, 2003). Such a practice can provide insights but is often used with the idea that such differences are natural and inherent, a perspective that has been heavily critiqued, as we discuss below. Rather, sociocultural context plays an overwhelming part in shaping gender differences, which vary across cultures and throughout time (Ceci et al., 2009). Parameterization, by focusing on a binary division, also ignores the multiple genders that are a part of many cultures (Lang, 1999; Nanda, 2014). A single individual’s gendered identity may also change significantly throughout their lives. Transgender persons are the most obvious example here; however, persons whose identities match their sex (cisgender persons) and who largely conform to gender norms may nevertheless shift throughout their lives in how they present themselves in ways that also challenge prevailing gender norms.

Towards Gender-Aware and Inclusive HCI

We outline various nuanced approaches to gender in HCI, distinguishing between gender-aware and gender-inclusive HCI. Gender awareness reflects on how gender may affect development practices and user interactions, explicitly recognizing the limitations of particular designs in terms of which genders or gendered characteristics are hindered and facilitated. Inclusivity actively seeks to include multiple and intersectional genders, and perhaps even future unknown users and characteristics. Regardless, the first step to awareness and inclusion is an understanding of the vocabulary and theories relating to gender developed in disciplines such as anthropology and gender studies. Being able to discuss gender explicitly and specifically works as a beginning towards understanding how gender relates to human interactions with technologies; further steps build on these vocabularies and theories. Yet, approaches to gender and technology are multiple, meaning these steps should not be taken as a simple linear process. Rather, we present tools and strategies for promoting gender diversity in HCI, reflecting on norms and values embedded in technological practices, and for understanding and accounting for these processes in all stages of HCI research and development.

Understanding gender

Popular usages of "gender" often equate it with "sex," such that there are two sexes (male and female) that are equivalent to two genders (masculine and feminine). This approach, referred to as essentialism, treats gender as an inherent and essential identity where gendered characteristics are natural and determined by various facets of biology. It is easy to see how such an approach can lead to stereotypes about "opposite" sexes/genders, where men and women have binary and opposing characteristics that can be applied in creating technologies.

Research from social science fields, however, has heavily critiqued this approach to gender (Butler, 1999; de Beauvoir, 1964; Haraway, 1991b; Mead, 1963). Anthropologist Margaret Mead (1963), for example, showed how behavior such as aggressiveness and cooperativeness, often seen as a natural part of being a man and woman, varied between sexes in three different groups in New Guinea. This clearly demonstrates how such masculine and feminine traits are not based on biological sex. Throughout history and across different groups and cultures, what it means to be a man or a woman varies significantly. It is therefore important to distinguish between sex and gender:

- **Sex**: A person's biological classification as male or female, including physical appearance, chromosomes, hormones, reproductive organs, and secondary sex characteristics.
- **Gender**: A person's self-identification and expression in relation to what is considered socially masculine, feminine, or other gendered category, within a given context.

It is also important to think beyond binary categories of man/woman and masculine/feminine given the growing legal recognition of third-gender categories in places such as Australia and India (Davidson, 2014; Mahapatra, 2014), as well as third gendered groups such as Hijra or Two-Spirited persons with a long history as part of Indian and some Indigenous cultures, respectively (Lang, 1999; Nanda, 2014). Additionally, there is significant variation within gender categories as they are currently used. While it is beyond the scope of discussion in this chapter, it is worth reflecting on the argument by some scholars who recognize more than two sexes, acknowledging different configurations of intersex persons and how the opposing binary of male and female has been socially constructed (Fausto-Sterling, 1993).

Theoretical approaches to gender Given a distinction between sex and gender, there are multiple ways gender can be approached, much of which has developed out of feminist and women's studies research. We briefly discuss some of the approaches that are key to doing gender-aware and gender-inclusive design. This discussion should be taken as an introduction—a small window into a vast body scholarship—that offers HCI researchers and practitioners some guidance and prompts them to explore further.

Early feminist activism and approaches in the United States and other Western countries focused on emancipation and achieving equal rights for women. These are often referred to as the first and second "waves" of feminism, and some of the issues and struggles, such as equal pay, are ongoing. For many of the issues, underlying these waves is a "liberal feminist" approach, which has been the approach to gender and feminism most commonly followed in HCI (Rode, 2011). This approach focuses primarily on differential opportunities accorded to men or women, with the idea that "'all other things being equal' gender should not matter" (Rode, 2011, p. 394). Such an approach, however, assumes that technologies are themselves gender neutral and ignores the complexities entailed in "all other things being equal" (Rode, 2011). Additionally, such a liberal feminist approach tends to treat a given gender as largely homogeneous—that issues facing women are all the same and equivalent. However, while there are likely some similar experiences shared among women (and among most men), there are also many points of difference.

"Third-wave" feminist theory has sought to acknowledge these differences, incorporating contributions and critiques from black feminists in the United States, and attempting to understand the lives and positions of women and feminists around the world. In particular, scholars argue for the need to explore how gender intersects with other facets of identity and systems of oppression, such as race, class, ethnicity, religion, and nationality, known as "intersectionality" (Crenshaw, 1989). Such an approach ties in with anthropological research that has shown global variation in masculinities and femininities, along with gendered categories (Mahmood, 2005; Mead, 1963; Nanda, 2014). It works to complicate liberal theories, pointing to ways some women (and men) may have social privileges not experienced by others.

One significant approach is feminist standpoint theory, which argues that women's lives and experiences produce different types of knowledge from men's, as part of their differing positions in the workings of power in patriarchal societies (Hartsock, 1983). Further expanded, such an approach suggests that persons with different social/cultural/economic positions hold different perspectives from which they see and approach the world, shaping what they do and know (Haraway, 1988). This approach has been critiqued for reifying differences between groups, such as men and women, as occurs with an essentialist perspective. However, it can be useful for highlighting the different approaches and perspectives women and other "marginal" groups might have in relation to various facets of technology design and use (Bardzell, 2010). We will also return to the significance of this perspective in discussing reflexivity.

Gender and queer theory also points to how gender is a practice—something that is *done*, rather than something that *is* (Butler, 1999; de Beauvoir, 1964). In particular, Judith Butler argues that gender is an ongoing and repetitive performance that refers to societal norms about gender. It is a continuous practice of dressing, moving, talking in ways seen as "normal" for a man or a woman. Sandra Harding's elucidation of layers through which gender can be explored is useful here. She suggests gender is simultaneously structural, symbolic, and individual (Harding, 1986). Structural gender entails the different social roles associated with particular genders, relating to divisions of labor, for example. Symbolic gender refers to the characteristics associated with a particular gendered role, such as the norms and behaviors relating to dress, hairstyle, or communication styles associated with being a man or woman. Individual gender is individual expression and self-identification as discussed above. In doing gender, a person's individual gender is therefore always performed in reference to symbolic and structural facets of gender. This approach and the challenges it brings to ideas of "normal" is known as queer theory, which has been used by LGBTQ movements for critiquing sexual normativity, but has much broader implications, also discussed below.

These are some of the key approaches that are important for understanding how gender is involved in HCI. There are many other perspectives, approaches, and literatures on gender, such as Marxist feminist theory (Haraway, 1991b), Islamic feminism (Alsheikh, Rode, & Lindley, 2011; Mahmood, 2005), and masculinity studies (Mellström, 2002), which could be useful for design and development in particular contexts. We encourage HCI practitioners to read widely about gender theories to understand the complexities and nuances to these theories.

Gender-aware HCI

With these theoretical tools for understanding gender, there are several ways to explore how gender is related to technology. Feminist technology studies scholars have shown how gender and technology are coproduced (Berg, 1996; Cockburn & Ormrod, 1993). In other words, doing gender is intimately related with interacting with technologies, and likewise how computers and technologies are designed, used, and operated is based on gendered norms, values, and behaviors. Working towards gender-aware technologies and interactions focuses on the latter direction of this relationship.

Scripts and metaphors One way to consider how gender norms and values becomes embedded into particular technologies and interactions is to consider gender "scripts" (Akrich, 1992). Such scripts relate to many facets of technological objects—their physical form, their placement, their interface, as well as marketing, advertisements, and instructional materials. Technologies are "inscribed" by the many people involved in product development and sale, which can be done accidentally or intentionally, and may also relate to other categories of difference such as socioeconomic status or race, along with gender.

The inscription of gender is clearly seen, for example, in the design and marketing of men's and women's shavers. Ellen van Oost (2003) shows how the men's shavers developed in the 1970s and 1980s projected and embodied particular ideas about gender—that men want to tinker and women prefer simplicity. By providing options for adjustability and incorporating monitoring features on electronic displays, men's shavers both repeated and reinforced the idea that men like and are good at using technology. In comparison, the women's shavers hid the technology by hiding screws and providing no options for adjustment. They were also marketed as a cosmetic product, suggesting shavers for women are not technological things and, in the process, women should not like or need to use technologies (van Oost, 2003). Brunner, Bennett, and Honey (1998) also explore the masculine gender scripts that are common in games, while also suggesting qualities, metaphors, and scripts that would incorporate open, diverse, and feminine characteristics as determined through qualitative research, approaches that are also discussed below in relation to gender-inclusive HCI.

When following a user-centered design methodology, designers may use personas to exemplify their users, giving them substance and nuance. Alternatively, designers might use scenarios creating narratives of technologies in use. These practices inherently and explicitly build ideas about how people will use the technology being designed, and who will be using them. Scripts (gendered or otherwise) are not inherently problematic. Yet, as seen with the shavers, there is a strong tendency to use stereotypes as part of personas or narratives about users, which can lead to inaccurate and exclusionary characterizations of users and their behaviors (Turner & Turner, 2011). As discussed, when done unconsciously, there is also a strong tendency for designers and developers to follow "I-methodology." Moreover, when the gender of users is left unspecified, developers will most likely create products that are the same as those created specifically for boys or men (Huff & Cooper, 1987).

The metaphors that are used to understand human interactions with technologies are also significant. Seeing computers and activities as akin to slaves or servants, for example, has gendered and racial connotations and focuses on dominance and control, frequently seen in relation to discourses about robots and AIs and their role in the future of humanity (for example Brahnam, Karanikas, & Weaver, 2011; Chasin, 1995). Alternatively, Donna Haraway has suggested using the metaphor of the cyborg, which is a "hybrid of machine and organism" entailing embodied, multiple, and partial identities (Haraway, 1991a, p. 149). Following such a metaphor would treat AI not as something poised to serve or overtake humanity, but as an interaction or relationship of play and possibility (Breslin, 2013). It is therefore worth thinking about what metaphors are used in creating designs and technologies and how metaphors of cyborgs versus servants change the interaction that is presumed and created.

Reflexivity For creating gender-aware designs, it is therefore important to be aware of the assumptions that designers and developers hold when making their products. Personas and other user-centered design techniques provide one way of explicitly defining who the technology is being designed for. Yet, implicit assumptions are likely still to be made in choosing what types of personas to include and what their characteristics are, as discussed above. A key step to creating gender-aware technologies, then, is reflexivity: self-awareness about one's own gender identity and values as they operate in practice. While objective measures—usability statistics, measures of number of clicks, eye tracking, and so on—are highly valued in HCI, it is clear that designers and developers are situated and embodied persons with particular cultural, religious, educational, and gendered lives (Forsythe, 2001). These particularities are why having more diverse teams *can* produce more diverse design perspectives. Yet, they are also simultaneously why a technologically savvy, university educated, white woman cannot represent the experiences of all women; likewise for a man with a similar background being able to represent all men (or all men and women).

As such, Donna Haraway (1988) has argued for using the notion of embodied vision—that perspectives are partial and seen from particular places and times—to explore the specificities of our own and "others'" embodied positions. The point is for HCI practitioners to be reflexive about their own embodied perspectives, to take account of the specificity of their knowledge and behavior, including the ways they are multiple, partial, and changing (Haraway, 1988, p. 583). Having a grasp on the vocabulary and theories for conceptualizing gender and other identities is one step towards such self-understanding. This embodied vision also includes accounting for the ways it is shaped by particular instruments for measurement and visualization (Haraway, 1988). For example, the ways in which statistical analyses divided by gender may emphasize differences between men and women, while hiding the heterogeneity within a category. This is not to say such analyses are not useful in certain circumstances, nor that their results are inaccurate or incorrect, but that who we are and the tools we use shape what we know and understand.

However, achieving such self-awareness is not a clear or straightforward process, nor is it likely fully attainable. Reflexivity should be an ongoing process of consideration, which can be aided by certain methodologies. Doris Allhutter (2012), for example, introduces "mind scripting" as one way for engineers and designers to reflect on their values and assumptions. Following this technique, participants produce anonymous text based on their memories of a particular event or topic, such as "the last time she or he tried a computer game" (Allhutter 2012, p. 693). These texts are then analyzed for the ways the author presents themselves, to reveal implicit ideas and assumptions. Mind scripting can then give insights into individual team members' gendered perspectives in relation to what constitutes a good computer game, what circumstances they are played in, and who likes computer games, for example.

Additionally, anthropological research has sought to expose and challenge naturalized values and assumptions by using interpersonal and participatory research, known as "participant observation." Margaret Mead's work discussed above is a classic example. Design research methods such as participatory design, which seeks to involve users democratically as part of the design process, provide one way of exploring users' perspectives and, when there is surprise or conflict, potentially exposing designers' own assumptions and values (van der Velden & Mortberg, 2014). Participatory design uses a wide variety of methods, including ethnography, along with future workshops, mock ups, storyboards, collaborative prototyping, seeking to understand the relationship between designers and others and following ethical practices (van der Velden & Mortberg, 2014). In general, methods that involve open-ended and rich interaction with users can work to challenge designers' and developers' inherent assumptions by confronting them with different perspectives, and to provide rich and detailed data about situated technology use (see also Sengers, Boehner, David, & Kaye, 2005). For gender-aware design, this reflexivity should be applied to design and development practice to produce awareness and accountability among those involved.

Designing accountably Gender-aware HCI is about accountability to those for whom the product is being designed for, and in the ways it is designed. We discuss in the next section the significance of broadly inclusive design. Yet, it is clear that, in some cases, technologies are meant for particular groups of people. Obvious examples might be breast pumps for women or an app that helps men recognize symptoms of prostate cancer. Even so, given the broad diversity of men and women, particular assumptions will be made that facilitate some users more than others. In the case of the breast pump, for example, what cultural values does it assume? Is it too big to carry with the assumption that women will always want to use it at home? Does it assume particular bodies? Such accountability is particularly significant when programs or technologies are meant for "everybody."

Building programs, technologies, and systems are not just the creation of particular objects, but are tied to practices of particular people in particular circumstances. Lucy Suchman (1994) shows how a project initially about automating clerical work was situated in a complex situation of management cost-cutting measures, contests over the definition of "knowledge work," and actual practices of clerical workers that did not match management's vision of mindless labor. Accountability brings into focus whose interests particular designs and projects serve. Work on HCI is then not about the creation of an objectively "good" product (because goodness is always situational), but rather "located accountability," whereby "design success rests on the extent and efficacy of one's analysis of specific environments of devices and working practices, finding a place for one's technology within them" (Suchman, 1994, p. 99). In other words, success is about how well a technology fits within a particular context.

Given this situatedness, any given design will likely exclude some group of persons at some point in time, as it is impossible to account for the full range of global and temporal diversity. Additionally, technologies are often used in unexpected ways as they are domesticated by users and adapted or repurposed to fit their lives and practices (Oudshoorn & Pinch, 2003). Designers cannot be expected to anticipate all future possibilities, although there is value in attempting to do so in order to debase assumptions and to open up possibilities for diversity, which we discuss below.

Maja van der Velden and Christina Mörtberg (2012) suggest considering how each iteration of a design creates particular "cuts," or multifaceted decisions about the design and its relationship to gender. The implication, however, is not that we can eliminate "problematic inscriptions of gender" but "based on our desire to design for the Other, which we will never be able to know fully" (van der Velden & Mortberg, 2012, p. 678). They argue that designers should be responsible for the "cuts" they make and how they affect others. By exploring the scripts and metaphors embedded in current technologies, and reflexively considering those being inscribed in human-technology interactions in the making, designers and developers create awareness and accountability for their "cuts." If these cuts reproduce norms and values that exclude, harm, or hinder particular groups, they should be accountable to that too.

Gender-inclusive design

Gender-aware designs focus on awareness and accountability relating what gendered (and other) values and norms are embedded in a design. Gender-inclusive designs go further and actively seek to incorporate feminist, culturally sensitive, diverse, and unexpected perspectives as part of design practice and the technologies that are produced (see also Breslin and Wadhwa 2014b). There are two overlapping facets to inclusive design: activist and intentionally inclusive approaches, and designing for future and unexpected diversity.

Feminist design In her key paper on Feminist HCI, Shaowen Bardzell (2010) discusses how, in serving existing needs HCI is often inherently conservative, acting to preserve the status quo in terms of norms, values, and social relations. She comments that if designers "are not careful they may perpetuate regressive and harmful practices and structures in service of usability" (Bardzell, 2010, p. 1306). Such norms and values often rely on stereotypes about men and women, as discussed, and can work to reinforce particular gender divisions or roles. For example, placing children's change tables only in women's washrooms reinforces the notion that women are responsible for childcare and blocks the possibility for men to contribute. Gender roles, norms, and distinctions currently do exist for many people, which are created and maintained through socialization, repeated discourses, and norms deeply embedded in various aspects of life, from infrastructures to census forms to building codes (Bowker & Star, 1999). Yet, people are not inherently tied to current or local definitions of what it means to be a man or woman. In other words, norms can be changed and one of the ways to do so is through feminist design.

Feminist HCI is one way to develop new designs and practices that are more inclusive towards varying and fluid genders. Several frameworks have been proposed as approaches to doing feminist design (Bardzell, 2010; Bath, 2014; Rode, 2011; van der Velden & Mortberg, 2012). Bardzell's approach rests on feminist standpoint theory, suggesting that attention to multiple standpoints can work as a way to incorporate "marginal" users as part of the design process. She points to several qualities that should be used as a constellation for a feminist approach to interaction design, namely pluralism, participation, advocacy, ecology, self-disclosure, and embodiment. Her discussion provides a general overview for what a feminist HCI could look like (Bardzell, 2010, p. 1305). Rode (2011) takes this as a basis, and, drawing on Harding's tripartite understanding of gender, works to provide a sociotechnical theory for gender and HCI based on three approaches: technology as masculine culture, gender positionality, and lived body experience.

A great deal of research has shown how technologies are often symbolically coded as masculine (Cockburn & Ormrod, 1993; Faulkner, 2001; Huff & Cooper, 1987). This association between masculinity and technology is prevalent globally, although it is not homogenous (e.g. Lagesen, 2008). Nevertheless, this association is further supported by the persistence of disparate numbers of women in technology fields in many places, leading to potential conflicts for women between femininity and working with technology. Rode therefore suggests incorporating feminine values as part of technologies and promoting the value of a technical femininity (Rode, 2011). This approach is different than designing for specifically for women, and instead entails creating technologies / interfaces for both men and women that appeal to characteristics often seen as feminine. In *some* cases, applying feminine stereotypes within products has the potential to transgress gender norms, as products may thereby contain multiple and sometimes contradictory representations of gender (Rommes, 2014, pp. 44–46). Yet, applying characteristics and practices of actual feminine users works against stereotypes *and* incorporates feminine practices and values.

Different values and practices associated with gender as it intersects with race, class, culture, and other categories of difference also need to be recognized and considered. This recognition is particularly important when focusing on HCI for broad use, such as infrastructures. In this regard, it also matters which approaches to gender and feminism are being applied in a given context. Alsheikh et al. (2011), for example, demonstrate the contrasting values that would be applied in designing VOIP and social media for usage in an Arabic cultural context when following liberal feminist versus Islamic feminist models. Liberal feminism, more familiar to Western HCI, places a high value on freedom and privacy and therefore could see practices such as women sharing their passwords with their husbands or choosing to veil as evidence of imbalances in power. On the other hand, from an Islamic feminist perspective, there is agency and power in the choice to follow Islamic practices and laws—in becoming a pious person (Alsheikh et al., 2011; Mahmood, 2005). The authors suggest designing technologies in ways that give women the choice of enacting piety or not. This provides a culturally and religiously relevant approach to gendered usages of these technologies, even if they may contrast with designers' values.

The other facets of Rode's (2011) sociotechnical theory provides some direction in accounting for cultural and other differences. She suggests looking to gender positionality as a way of understanding changing gender roles in relation to technology, including ways that do not fit with a binary categorization of gender. The significance here is to explore how individuals relate to gender norms, but also how those norms are contested or simply irrelevant as part of individual lives. An attention to lived body experience also points to the complexities of life as they are lived, concomitant with the messiness of identities and situations (Rode, 2011). This works against assumptions about how a person uses a particular technology according to their categorical situation—being a woman, being white, being poor—to focus on how users (people) use technologies in action. These three approaches—technology as masculine culture, gender positionality, and lived body experience—provide a useful way of applying a nuanced understanding of gender. In focusing on the intersectionality, positionality, and experience of gender and of particular lives, it becomes less tenable to rely simply on ideas that women like pink or men like war games. Of course, such characteristics are true of some men and women, but a goal of feminist design is to allow freedom of choice and difference unconstrained by stereotypes or restricted conceptualizations of femininity and masculinity.

However, these theoretical approaches only provide limited guidance on concrete methods to follow. There is much research still to be done in this area. Most generally, gender inclusivity needs to be actively embedded in all facets of an organization: it needs to be a value understood and shared by management, researchers, and team members, and needs to be integrated in meaningful ways in all phases of design and development. We have suggested elsewhere to integrate gender HCI as part of a computing and HCI curriculum, as one way to provide students and professionals with knowledge about gender theories (Breslin & Wadhwa, 2014b, 2015). Qualitative and participatory methods in general also fit well with attention to user experiences and identities. Value-sensitive design, which seeks to uncover and define values throughout the design process, using a combination of conceptual, empirical, and technical investigations, has many shared goals with gender-inclusive design and therefore can point to useful methodological practices (Friedman & Kahn, 2002). Corinna Bath also argues for following a cyclic process of test and development from the start of the design process, as well as situating such user research in the context of social science research to evaluate values and perspectives, as users themselves can also perpetuate "traditional" or stereotyped gender norms (Bath, 2014). Thus, a combination of theoretical/conceptual frameworks and empirical user-centered research methodologies can be ways of working towards and doing feminist design.

Queer and future-oriented design A challenge of feminist and value-sensitive design is that users and designers can hold conflicting values. Drawing on queer theory, Ann Light (2011) suggests, rather than trying to "social engineer" particular values, designs should allow for flexibility and diversity. As Cassell (2003) comments in relation to her approach of "undetermined design": "we didn't see that it was our place to claim to know what a girl was or what a boy was, because there's too much diversity." While "flexibility" and "diversity" are values in themselves, the point is to open up possibilities—to allow for indeterminacy—for current and future use by a wide variety of unexpected persons, rather than to restrict or determine how a program or technology should be used.

Undetermined design is one strategy where users are given interpretive flexibility, to "attribute themselves a gendered identity… to create or perform themselves through using technology" (Cassell, 2003, p. 13). It seeks to work in ways such that design continues throughout the process of technology use. A very basic example is Facebook's gender options, seen in Figures 4.1 and 4.2. While they include the binary female and male, they also include a choice for "Custom." Typing brings up a list of common options, such as "trans" or "gender questioning," yet users can also type their own gender. However, the cultural specificity of Facebook's development becomes evident here as Two-Spirit, part of some Indigenous cultures, is included in

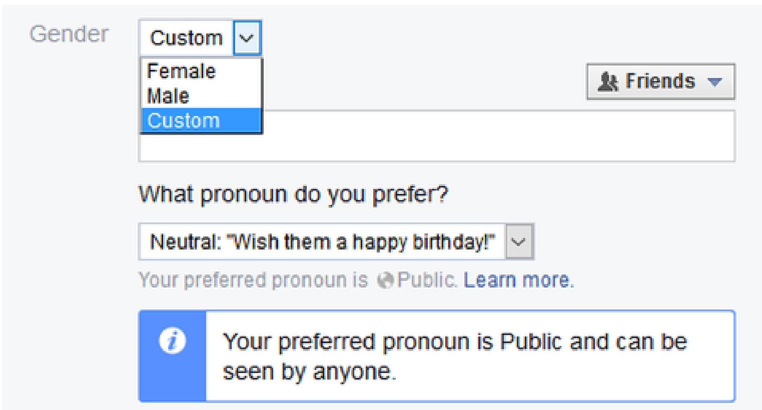


Figure 4.1 Facebook gender options.

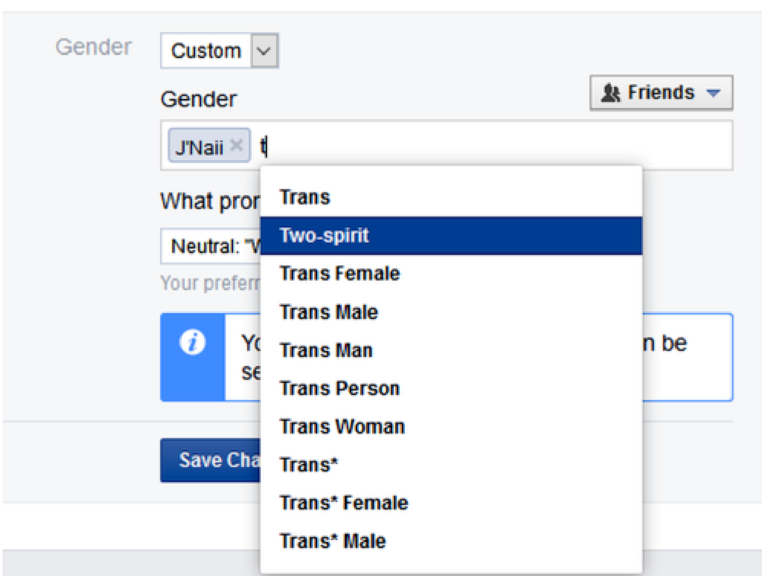


Figure 4.2 Facebook custom gender.

the form-fill list, but not Hijra who are a third gender group in India with a long history and recent legal recognition (Mahapatra, 2014; Nanda, 2014). The inclusion of male and female as default options, as well as the autofill features therefore reinforce certain genders as more common—more "normal"—than others. Yet, the undetermined option for writing a user-defined gender allows users to create and perform their gendered identities through their Facebook accounts, should they wish.

Possibilities for design can be much more extensive, such as stuffed animals that work as children's allies and friends, encouraging the child to create and tell stories that are recorded and edited along the way, and so allowing children to learn and create themselves through their stories (Cassell, 2003). The significance of indeterminacy—what can be seen as "future-oriented" design—is to engage users not as passive consumers of objects, but for them to be cocreators in their own identities and in the technologies themselves. Values or norms are therefore unstated, as they are left open for the users to perform and create, or to disrupt and challenge.

Similarly seeking to open up possibilities as part of design, Ann Light points to several ways that systems can be made less determinate. She draws on Judith Butler's (1999) work and the concept of "queering," which attempts to bring to the forefront seemingly natural but nevertheless arbitrary norms, and ways of disrupting or "troubling" them. Forgetting, obscuring, cheating, eluding, and designing for obliqueness are particular ways of queering. Light points out there are certain cases where queering is inappropriate—in general, we want controls for airplanes or ovens to be straightforward. The goal, however, is to challenge designers "to consider not just how to design for the variation in culture that we find worldwide, across gender, race, ethnicity and so forth, but to design so that we do as little as possible to hamper the evolution of variety" (Light, 2011, p. 436).

This focus on diversity and inclusivity finds a counterpart in Universal Design / Design for All, which focused initially on accessibility to technological and physical infrastructure, but more broadly seeks to recognize diversity in embodiments and contexts of use (Stephanidis, 2001). Gender-inclusive and Design for All seeks to facilitate everyone's use of technologies in the way that is captured by the best science fiction, which questions what it is to be human, as much as a man or a woman, and reflects on those possibilities towards more aware and inclusive practices.

Conclusion

Gender is already an integrated facet of HCI, but its role has often gone unacknowledged. Gender plays a part in the numerical disparity between men and women in computing education and careers, as well as in how technologies have been designed and developed. Through the I-methodology, stereotyping, and focusing on gender differences, gendered ideas of who users are, and what it means to be a man or woman, are being embedded in technologies and interfaces. We have discussed practices and approaches that are key to explicitly understanding and addressing the role of gender in HCI. We distinguish between gender-aware and gender-inclusive design. The former analyzes and accounts for the gendering incorporated in particular designs and practices, whereas inclusive designs actively seek to include diverse gendered identities, and allow for flexibility and self-definition.

The first step to creating gender-sensitive designs (aware and inclusive) is to understand gender—the vocabularies and theories about gender and how it relates to identities, experiences, and knowledges. Building on these understandings, we discuss practices for gender-aware and inclusive design. As we have previously highlighted, activism, multiple and intersectional perspectives on identity, and reflexivity are key approaches for gender HCI (Breslin & Wadhwa, 2014a, 2014b). Gender awareness entails analysis of the gender scripts and metaphors being inscribed into technologies, reflexivity on practitioners' and researchers' own gendered perspectives, and accountability for how the designs and technologies being made affect current and future users. Gender inclusivity focuses on recognizing and incorporating multiple and diverse gender values and norms into a unified project and allowing for identity creation through technology use and for the broadest possible gender diversity for current and future users.

References

Akrich, M. (1992). The de-scription of technological objects. In W. E. Bijker & J. Law (Eds.), *Shaping technology/Building society: Studies in sociotechnical change* (pp. 205–224). Cambridge, MA: MIT Press.

Allhutter, D. (2012). Mind scripting: A method for deconstructive design. *Science, Technology and Human Values*, *37*(6), 684–707. Retrieved from http://doi.org/10.1177/0162243911401633

Alsheikh, T., Rode, J., & Lindley, S. (2011). (Whose) value-sensitive design: A study of long-distance relationships in an Arabic cultural context. In *Proceedings of the 2011 ACM Conference on Computer Supported Cooperative Work (CSCW 2011)* (pp. 75–84). New York, NY: ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1958836

Bardzell, S. (2010). Feminist HCI: taking stock and outlining an agenda for design. In *CHI'10 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1301–1310). New York, NY: ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1753521

Barker, L., Mancha, C., & Ashcraft, C. (2014). *What is the impact of gender diversity on technology business performance? Research summary.* Retrieved from http://www.ncwit.org/sites/default/files/resources/impactgenderdiversitytechbusinessperformance_print.pdf

Bath, C. (2014). Searching for methodology: Feminist technology design in computer science. In W. Ernst & I. Horwath (Eds.), *Gender in science and technology: Interdisciplinary approaches.* Bielefeld: transcript Verlag.

Beckwith, L., & Burnett, M. (2004). Gender: An important factor in end-user programming environments? In *Proceedings of the 2004 IEEE Symposium on Visual Languages and Human Centric Computing (VLHCC'04)* (pp. 107–114). Rome: IEEE. Retrieved from http://doi.org/10.1109/VLHCC.2004.28

Berg, A. (1996). *Digital feminism.* Trondheim: Norwegian University of Science and Technology.

Berg, A. (1999). A gendered socio-technical construction: The smart house. In D. Mackenzie & J. Wajcman (Eds.), *The social shaping of technology* (pp. 301–313). Buckingham, UK: Open University Press.

Bowker, G. C., & Star, S. L. (1999). *Sorting things out: Classification and its consequences.* Cambridge, MA: MIT Press.

Brahnam, S., Karanikas, M., & Weaver, M. (2011). (Un)dressing the interface: Exposing the foundational HCI metaphor "computer is woman." *Interacting with Computers*, *23*(5), 401–412. Retrieved from http://doi.org/10.1016/j.intcom.2011.03.008

Breslin, S. (2013). 01010000 01001100 01000001 01011001: Play elements in computer programming. *American Journal of Play*, *5*(3), 357–382.

Breslin, S., & Wadhwa, B. (2014a). EnGendering interaction design. *Third International Conference on User Science and Engineering (i-USEr)* (pp. 292–295). New York, NY: IEEE. Retrieved from http://doi.org/10.1109/IUSER.2014.7002719

Breslin, S., & Wadhwa, B. (2014b). Exploring nuanced gender perspectives within the HCI community. *Proceedings of the India HCI 2014 Conference on Human Computer Interaction—IHCI '14* (pp. 45–54). New York, NY: ACM. Retrieved from http://doi.org/10.1145/2676702.2676709

Breslin, S., & Wadhwa, B. (2015). Towards a gender HCI curriculum. In *CHI'15 extended abstracts on human factors in computing systems* (pp. 1091–1096). Seoul, South Korea: ACM.

Brunner, C., Bennett, D., & Honey, M. (1998). Girl games and technological desire. In J. Cassell & H. Jenkins (Eds.), *From Barbie to Mortal Kombat: Gender and computer games* (pp. 72—88). Cambridge, MA: MIT Press.

Butler, J. (1999). *Gender trouble: Feminism and the subversion of identity* (10th anniv). New York: Routledge.

Cassell, J. (2003). Genderizing human-computer interaction. In J. Jacko & A. Sears (Eds.), *The human-computer interaction handbook* (pp. 401–412). Hillsdale, NJ: Laurence Erlbaum Associates Inc.

Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin*, *135*(2), 218–261. Retrieved from http://doi.org/10.1037/a0014412

Chasin, A. (1995). Class and its close relations: Identities among women, servants, and machines. In J. Halberstram & I. Livingston (Eds.), *Posthuman bodies* (pp. 73–96). Bloomington, IN: Indiana University Press.

Churchill, E. F. (2010). Sugared puppy-dog tails: Gender and design. *Interactions*, *17*(2), 52–56.

Cockburn, C., & Ormrod, S. (1993). *Gender and technology in the making*. Thousand Oaks, CA: Sage.

Cooper, J., & Kugler, M. B. (2009). The digital divide: The role of gender in human–computer interaction. In A. Sears & J. A. Jacko (Eds.), *Human-computer interaction: Designing for diverse users and domains* (pp. 3–15). Boca Raton, FL: CRC Press.

Crenshaw, K. (1989). Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. *University of Chicago Legal Forum*, *1989*(1), 139–167.

Davidson, H. (2014, April 2). Third gender must be recognized by NSW after Norrie wins legal battle. *Guardian*. Retrieved from http://www.theguardian.com/world/2014/apr/02/third-gender-must-be-recognised-by-nsw-after-norrie-wins-legal-battle

De Beauvoir, S. (1964). *The second sex*. (H. M. Parshley, Ed.). New York, NY: Knopf.

Faulkner, W. (2001). The technology question in feminism: A view from feminist technology studies. *Women's Studies International Forum*, *24*(1), 79–95.

Fausto-Sterling, A. (1993, March/April). The five sexes: Why male and female are not enough. *The Sciences*, pp. 20–25.

Forsythe, D. (2001). *Studying those who study us: An anthropologist in the world of artificial intelligence*. Stanford, CA: Stanford University Press.

Friedman, B., Kahn Jr., P. H., & Borning, A. (2002). *Value sensitive design: Theory and methods* (UW CSE Technical Report 02-12-01). Retrieved from http://faculty.washington.edu/pkahn/articles/vsd-theory-methods-tr.pdf

Haraway, D. J. (1988). Situated knowledges: The science question in feminism and the privilege of partial perspective. *Feminist Studies*, *14*(3), 575–599.

Haraway, D. J. (1991a). A cyborg manifesto: Science, technology, and socialist-feminism in the late twentieth century. In *Simians, cyborgs and women: The reinvention of nature* (pp. 149–182). New York, NY: Routledge.

Haraway, D. J. (1991b). "Gender" for a Marxist dictionary: The sexual politics of a word. In *Simians, cyborgs, and women: The reinvention of nature* (pp. 127–148). New York, NY: Routledge.

Harding, S. (1986). *The Science question in feminism*. Ithaca, NY: Cornell University Press.

Hartsock, N. C. M. (1983). The feminist standpoint: Developing the ground for a specifically feminist historical materialism. In S. Harding & M. B. Hintikka (Eds.), *Discovering reality* (pp. 283–310). Dordrecht: D. Reidel.

Huff, C., & Cooper, J. (1987). Sex bias in educational software: The effect of designers' stereotypes on the software they design. *Journal of Applied Social Psychology*, *17*(6), 519–532.

Lagesen, V. A. (2008). A cyberfeminist utopia? Perceptions of gender and computer science among Malaysian women computer science students and faculty. *Science, Technology and Human Values*, *33*(1), 5–27. Retrieved from http://doi.org/10.1177/0162243907306192

Lang, S. (1999). Lesbians, men-women and two-spirits: homosexuality and gender in native American cultures. In E. Blackwood & S. Wierenga (Eds.), *Feminist desires* (pp. 91–116). New York, NY: Columbia University Press.

Light, A. (2011). HCI as heterodoxy: Technologies of identity and the queering of interaction with computers. *Interacting with Computers*, *23*(5), 430–438. Retrieved from http://doi.org/10.1016/j.intcom.2011.02.002

Mahapatra, D. (2014, April 15). Supreme Court recognises transgenders as third gender. *The Times of India*. Retrieved from http://timesofindia.indiatimes.com/india/Supreme-Court-recognizes-transgenders-as-third-gender/articleshow/33767900.cms

Mahmood, S. (2005). *Politics of piety: The Islamic revival and the feminist subject*. Princeton, NJ: Princeton University Press.

Margolis, J., & Fisher, A. (2002). *Unlocking the clubhouse: Women in computing*. Cambridge, MA: MIT Press.

Mead, M. (1963). *Sex and temperament in three primitive societies*. New York, NY: Morrow.

Mellström, U. (2002). Patriarchal Machines and Masculine Embodiment. *Science, Technology and Human Values*, *27*(4), 460–478. Retrieved from http://doi.org/10.1177/016224302236177

Nanda, S. (2014). *Gender diversity: Crosscultural variations*. Long Grove, IL: Waveland Press.

Oudshoorn, N., & Pinch, T. J. (Eds.). (2003). *How users and non-users matter: The co-construction of users and technology. How users matter—The co-construction of users and technology*. Cambridge, MA: MIT Press.

Oudshoorn, N., Rommes, E., & Stienstra, M. (2004). Configuring the user as everybody: Gender and design cultures in information and communication technologies. *Science, Technology, and Human Values*, *29*(1), 30–63. Retrieved from http://doi.org/10.1177/0162243903259190

Quinn, M. (2014, December 22). Quinn : Hey Apple Health, did you forget about women ? *San Jose Mercury News*. Retrieved from http://www.mercurynews.com/michelle-quinn/ci_27173258/quinn-hey-apple-health-did-you-forget-about?source=infinite

Rode, J. A. (2011). A theoretical agenda for feminist HCI. *Interacting with Computers*, *23*(5), 393–400. Retrieved from http://doi.org/10.1016/j.intcom.2011.04.005

Rommes, E. (2014). Feminist interventions in the design process. In W. Ernst & I. Horwath (Eds.), *Gender in science and technology: Interdisciplinary approaches* (pp. 41–57). Bielefeld: transcript Verlag.

Sengers, P., Boehner, K., David, S., & Kaye, J. "Jofish." (2005). Reflective design. *Proceedings of the 4th Decennial Conference on Critical Computing between Sense and Sensibility—CC '05* (p. 49). New York, NY: ACM. Available at http://doi.org/10.1145/1094562.1094569

Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, *35*(1), 4–28. Retrieved from http://doi.org/10.1006/jesp.1998.1373

Stephanidis, C. (2001). *User interfaces for all: Concepts, methods, and tools.* Portland, OR: Lawrence Erlbaum.

Suchman, L. (1994). Working relations of technology production and use. *Computer Supported Cooperative Work (CSCW)*, *2*, 21–39.

Tan, D. S., Czerwinski, M., & Robertson, G. (2003). Women go with the (optical) flow. In *Proceedings of the Conference on Human Factors in Computing Systems—CHI '03*, *5,* 209. Retrieved from http://doi.org/10.1145/642647.642649

Turkle, S. (1988). Computational reticence: Why women fear the intimate machine. In C. Kramarae (Ed.), *Technology and women's voices: Keeping in touch* (pp. 33–49). New York, NY: Routledge.

Turner, P., & Turner, S. (2011). Is stereotyping inevitable when designing with personas? *Design Studies*, *32*(1), 30–44. Retrieved from http://doi.org/10.1016/j.destud.2010.06.002

Van der Velden, M., & Mortberg, C. (2012). Between need and desire: Exploring strategies for gendering design. *Science, Technology and Human Values*, *37*(6), 663–683. Retrieved from http://doi.org/10.1177/0162243911401632

Van der Velden, M., & Mortberg, C. (2014). Participatory design and design for values. In J. van den Hoven, P. E. Vermaas & I. van der Poel (Eds.), *Handbook of ethics, values, and technological design* (pp. 1–22). Dordrecht: Springer Science and Business Media. Retrieved from http://doi.org/10.1007/978-94-007-6994-6

Van Huysse, J. (2014, July 23). Building a Twitter we can be proud of. *Twitter Blog*. Retrieved from https://blog.twitter.com/2014/building-a-twitter-we-can-be-proud-of

Van Oost, E. (2003). Materialized gender: How shavers configure the users' femininity and masculinity. In N. Oudshoorn & T. Pinch (Eds.), *How users matter: The co-construction of users and technologies* (pp. 193–208). Cambridge, MA: MIT Press.

Williams, M. (2014, June 25). Building a more diverse Facebook. *Facebook Newsroom*. Retrieved from http://newsroom.fb.com/news/2014/06/building-a-more-diverse-facebook/

5

Usability and Digital Typography

Peter Flynn

Type design and typography have always been about usability: an unreadable document is not usable. Digital type brought a new flexibility to print and screen design, but the full effects of this have yet to be worked through. The fluid circumstances in which digital type is used mean that we need to remain aware of how applying simple rules in the appropriate circumstances can be used to improve the contribution of digital type to human-computer interaction. An understanding of usability and measurement can help improve the quality of documents, and also reveal where there is still scope for more work.

Digital type, one of the great achievements of the digital world, may have had unintended consequences.

Most of us who use computers work in a graphical computing environment, whether it's OS X, the X Window system, Microsoft Windows, Android, or iOS. Every app we use, every browser, every desktop, and every application window uses fonts—sometimes lots of them. Even people who work in nongraphical console environments like Terminal, Xterm, or Command Prompt will see at least one font—typewriter-style fixed-width characters. And all the characters we read in a document created on a computer have been assembled by a font-using program of some kind. Figure 5.1 shows the same information in three applications using entirely different fonts.

The great achievement is that someone, somewhere, at some point in the last 30 or so years created the patterns of tiny dots that you see (or more likely, don't see) in front of you right now, regardless of whether you're reading this from paper or from a screen. Possibly it was many people, in many places; and more likely what they created was the programs that create the dots, rather than the dots themselves (computer fonts are actually programs for drawing characters). We manipulate these characters in our word processors or editors whenever we type, copy, paste, or delete them. We may also be able to resize them, invert, mirror, flip, rotate, distort, color, and even decorate them. This was mostly impossible in earlier centuries when fonts were made of metal types, cast in tiny molds (the word "font" comes from the French for "melted").

The Wiley Handbook of Human Computer Interaction, Volume 1, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

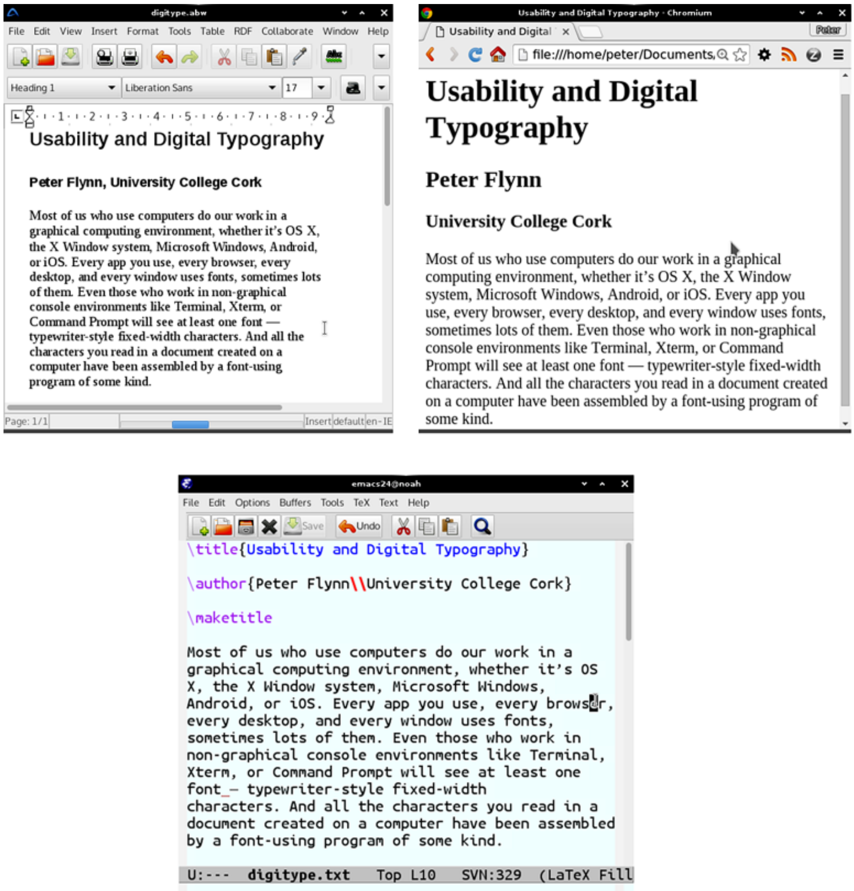


Figure 5.1 Different fonts in a word processor (AbiWord), a browser (Chrome), and an editor (Emacs), but the same text. *Source*: CC:BY-NC-SA © 2016 Author.

If you're reading this in Braille, the pattern of dots was created in 1824, which was an even earlier achievement. Typesetting with "hot metal," where the types are cast and set all in one operation, was developed in the 1880s–1890s.

The unintended consequence is that everyone's doing it. Typesetting used to be done exclusively by compositors who had undergone a traditional 7-year craft apprenticeship. They "came out of their time" at least knowing their craft thoroughly—as do professional compositors today—even if they weren't actually designers. But decades of text editors, word processors, desktop publishing systems, and wildly overenthusiastic marketing have led us to believe that nowadays we can all design and typeset everything ourselves, with results just as good as the professionals.

Unfortunately this isn't always the case: we are surrounded by typographic errors and design faults that proper training would have made impossible before. Fortunately, most people reading what we typeset don't notice when there are flaws. And happily, many people have learned from the accumulated knowledge and skill of the previous

500 years (often the hard way, by making mistakes). However, unless we pay for the professionals, most of us just have to do the best we can with what skills we've got. The trick is to know what to look out for.

As we have to use some of the vocabulary of the Web, graphic, and printing industries, there is a short glossary of technical terms towards the end of this chapter to try to clear up any ambiguities. I have avoided using some of the more abstruse terms of the art, but you may encounter them in other writing on the subject.

Human-computer Interaction (HCI), Usability, and Typography

So what does all this have to do with HCI or usability?

Type and printing have had a long relationship with usability, starting with type design—Gutenberg created the font for his famous Bible in the 1450s (see Figure 5.2) to resemble the writing of the scribes who produced books up until then. In effect he was trying to make his printed Bible as usable as the handwritten ones everyone literate already knew (Loxley, 2005, p. 13). In the history of type design, it is clear that type designers have always wanted to make their typefaces usable (that is, aesthetically pleasing and appropriate; a "good user experience") for specific purposes, whether it was a series of popular classics in pocket editions (Griffo's italic type for Aldus Manutius), a Greek font for the French royal press (Garamond), a new face for a newspaper (Times)—or a digital font for typesetting mathematics (Computer Modern).

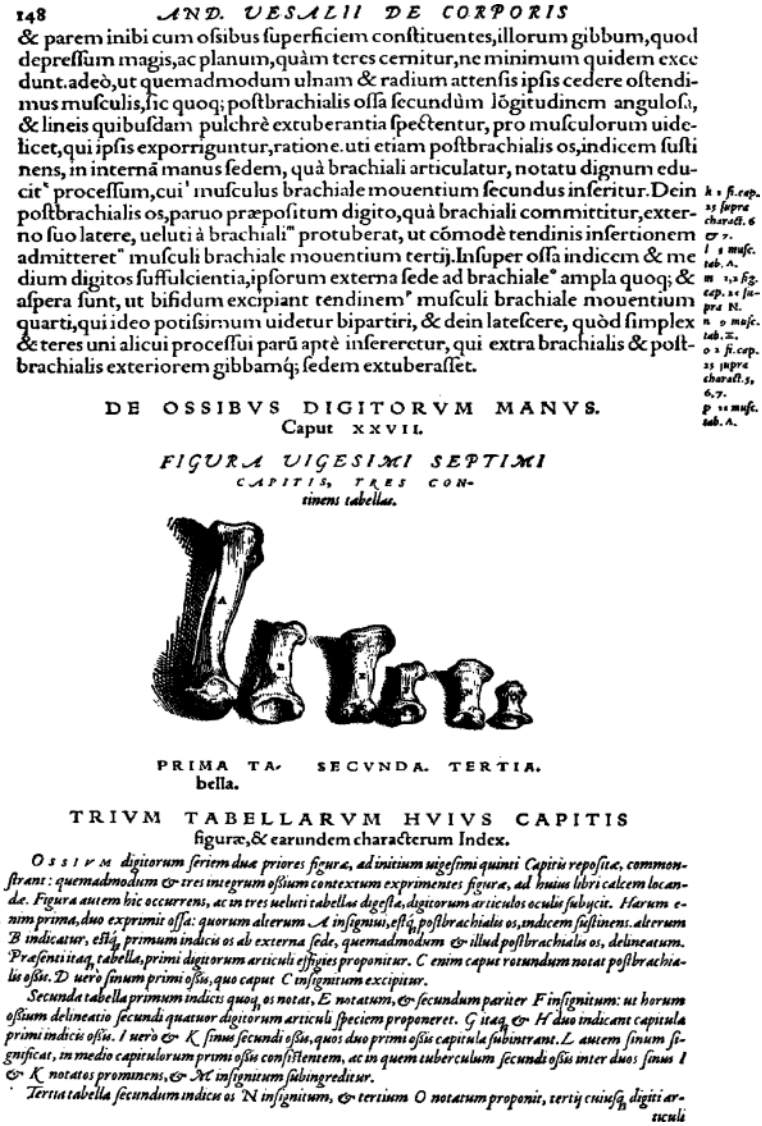
A similar case can be made for page design or layout. Documents in the west, especially books, have a much longer history than type does. They developed from the incised clay tablets of 5,000 years ago, through the folded wooden or papyrus books, and scrolls of the Mediterranean civilizations, through the codex, or hardcover book of handwritten pages that appeared in Europe around 500CE, right through the age of printing to modern paperbacks, digital editions, and eBooks. On that journey, documents have evolved to share a common set of ways to convey different types of information: chapters and verses, titles and subtitles, headings and subheadings, paragraphs, lists of all kinds, tables, figures, notes, indexes, and many others.

These "elements of information" acquired broadly standardized **meanings**, although **appearances**, of course, varied. At times the design of a page was even apparently made deliberately hard to understand—legal documents of the Middle Ages required highly specialist training just to read (Tiersma, 2010, pp. 150–151), let alone understand; possibly in part to ensure that you continued to need a lawyer! But by only a century after Gutenberg, we already see pages that are recognizably modern, with pages almost indistinguishable from a present-day textbook (Figure 5.3) and we also find authors and printers using different layouts in an attempt to present the information as usably as they saw fit. In following the patterns of established use and adapting the presentation to suit the readership, designers were trying to make the result usable.

The conclusion has to be that type designers and typographic designers (and Vesalius saw himself as much a designer as an author (Saunders & O'Malley, 1950, p. 46)) have

Figure 5.2 First page of the Epistle of St. Jerome (his introduction to the Bible) from the University of Texas copy of a Gutenberg Bible. The page has 40 lines, meaning it came from one of the earliest folios, before the line-count was increased to 42. *Source:* Public domain by courtesy of Ransom Center, University of Texas at Austin (http://www.hrc.utexas.edu/exhibitions/permanent/gutenberg/); for further information, see https://en.wikipedia.org/wiki/Gutenberg_Bible#/media/File:Gutenberg_bible_Old_Testament_Epistle_of_St_Jerome.jpg

probably been doing for a long time much what they do today, and as we do in HCI: designing with a defined target population and a use-case in mind. This is fundamentally a user-centered approach—it has to be, because if a publisher produced a book or newspaper that no one could read, no one would buy it.


148 AND. VESALII DE CORPORIS

& parem inibi cum oſsibus ſuperficiem conſtituentes, illorum gibbum, quod depreſſum magis, ac planum, quàm teres cernitur, ne minimum quidem excedunt. adeò, ut quemadmodum ulnam & radium attenſis ipſis cedere oſtendimus muſculis, ſic quoq; poſtbrachialis oſſa ſecundùm lõgitudinem anguloſa, & lineis quibuſdam pulchrè extuberantia ſpectentur, pro muſculorum uidelicet, qui ipſis exporriguntur, ratione. uti etiam poſtbrachialis os, indicem ſuſtinens, in internã manus ſedem, quà brachiali articulatur, notatu dignum educit[k] proceſſum, cui[l] muſculus brachiale mouentium ſecundus inſeritur. Dein poſtbrachialis os, paruo præpoſitum digito, quà brachiali committitur, externo ſuo latere, ueluti à brachiali[m] protuberat, ut cõmodè tendinis inſertionem admitteret[n] muſculi brachiale mouentium tertij. Inſuper oſſa indicem & medium digitos ſuffulcientia, ipſorum externa ſede ad brachiale[o] ampla quoq; & aſpera ſunt, ut bifidum excipiant tendinem[p] muſculi brachiale mouentium quarti, qui ideo potiſsimum uidetur bipartiri, & dein lateſcere, quòd ſimplex & teres uni alicui proceſſui parũ aptè inſereretur, qui extra brachialis & poſtbrachialis exteriorem gibbamq; ſedem extuberaſſet.

k 1 fi. cap. 25 ſupra charact. 6 & 7.
l 5 muſc. tab. A.
m 1, 2 fig. cap. 25 ſupra N.
n 9 muſc. tab. X.
o 2 fi. cap. 25 ſupra charact. 5, 6, 7.
p 11 muſc. tab. A.

DE OSSIBVS DIGITORVM MANVS.
Caput XXVII.

FIGVRA VIGESIMI SEPTIMI CAPITIS, TRES CONtinens tabellas.

PRIMA TAbella. SECVNDA. TERTIA.

TRIVM TABELLARVM HVIVS CAPITIS figuræ, & earundem characterum Index.

OSSIVM digitorum ſeriem duæ priores figuræ, ad initium uigeſimi quinti Capitis repoſitæ, commonſtrant: quemadmodum & tres integrum oſsium contextum exprimentes figuræ, ad huius libri calcem locandæ. Figura autem hic occurrens, ac in tres ueluti tabellas digeſta, digitorum articulos oculis ſubijcit. Harum e-
A. nim prima, duo exprimit oſſa: quorum alterum A inſignitum, eſtq; poſtbrachialis os, indicem ſuſtinens. alterum
B. B indicatur, eſtq; primum indicis os ab externa ſede, quemadmodum & illud poſtbrachialis os, delineatum.
C. Præſenti itaq; tabella, primi digitorum articuli effigies proponitur. C enim caput rotundum notat poſtbrachia-
D. lis oſsis. D uerò ſinum primi oſsis, quo caput C inſignitum excipitur.
E.F. Secunda tabella primum indicis quoq; os notat, E notatum, & ſecundum pariter F inſignitum: ut horum
G,H. oſsium delineatio ſecundi quatuor digitorum articuli ſpeciem proponeret. G itaq; & H duo indicant capitula
I,K. primi indicis oſsis. I uerò & K ſinus ſecundi oſsis, quos duo primi oſsis capitula ſubintrant. L autem ſinum ſi-
L. gnificat, in medio capitulorum primi oſsis conſiſtentem, ac in quem tuberculum ſecundi oſsis inter duos ſinus I
M. & K notatos prominens, & M inſignitum ſubingreditur.
N.O. Tertia tabella ſecundum indicis os N inſignitum, & tertium O notatum proponit, tertij cuiuſq; digiti articuli

Figure 5.3 A page from the medical textbook *De Humani Corporis Fabrica* (*On the Fabric of the Human Body*) of Vesalius, 1543 edition, showing the recognizable features of a modern text. *Source:* Copyright © 2009 Google Books.

This remains true in digital form, but we have an escape route that was unavailable in the predigital age. A badly designed, broken, or unreadable Web page, word processor file, or PDF document can at least be enlarged onscreen, or perhaps have the styling disabled, or be read aloud by a screen reader, or even be copied and pasted in plaintext form with all the formatting stripped out (assuming it's not copy protected). Then the text can be read, regardless of how bad the original was. This doesn't excuse poor design or sloppy typography but it does highlight that the usability of digital type can operate at several different levels.

Can We Measure the Usability of Digital Type?

If we are right about the usability intentions of type designers and typographic designers—digital or not—then whatever degree of usability we see, it is built in from the start. Books and type, even digital ones, tend to be cast in concrete at the time of publishing: the nature of their commonest forms (paper, pdf, eBook) rather precludes easy alteration afterwards. Although it is technically possible with digital forms to reformat them, it requires specialist knowledge and tools, and may in any event be prohibited by copyright or disabled by digital rights management (DRM) software. In practice, only HTML publication on the Web remains open to complete reformatting and easy updating.

That being so, we have to consider measurement at the point of creation, where the decisions on format were made. There are many important aspects to a document, not least of which is the quality of the writing itself, but for typographic purposes it is useful to divide the act of creating a document into three broad areas:

- type design and choice of font;
- typographic design and layout;
- the process of composition.

By "document" here we mean any surface, physical or logical, on which we put text for others to read. The weight given to each of the three areas will of course vary between types of document; for example, a set of road signs or instrumentation labels will have very different requirements from a concert poster, a book on HCI, or the information leaflet in a pharmaceutical product, but this classification lets us explain some of the common features that affect usability.

Type design

Type design means the shape of the letters and their spacing. How tall should the "d" be compared with the "x" (remember the Dunhill cigarette packet font)? How round should the "e" be, and should the crossbar be horizontal or slanted? Should the letters have serifs (tails) or not? An elegant, classical design, or robust and authoritative, or grunge or fun? The fine detail of these questions is out of scope for most of us because we can't usually design typefaces ourselves—but they do affect one of the most important design decisions of all: the choice of typeface (see Figure 5.4).

The range of typefaces is vast, and it's tempting to be led into believing that certain fonts are "better" for certain uses or topics than others. The typographic

Computer Modern	Times New Roman	Univers	Punk
CM Dunhill	Centaur	Helvetica	Lobster Two

Figure 5.4 Some typefaces to match the design decisions in Type Design (all at 24pt). Image: CC-BY-NC-SA © 2016 Author.

designer John Lewis said he was once asked to design examples illustrating this: delicate little script fonts for cosmetic adverts; classical, formal, respectable roman faces for banks; big, chunky sans-serif fonts for engineering, and so on; only to discover, after a while, that he could "change the typefaces around at will and with ever increasing effect" (Lewis, 1963, p. 52).

Measuring the usability of a font has to take into account both perception (can the reader actually see the letters?) and cognition (do they resolve into words whose meaning can be grasped?)—more on this in the section on vision and usability below. At the optical level, letters need to be big enough to be seen (there is a reason why contracts use small print!) and they generally need to be separated by enough white space for accurate perception to occur. This spacing, known as **kerning**, is built into the characters in the font, and is normally automatic, but designers of display material (signage) sometimes adjust the kerning to close up or space out the gaps for aesthetic reasons. There are many resulting instances of the misrecognition of adjoining letters that were actually touching, or so close that they were seen as a single letter. The cynical and usually salacious examples of "keming" (a jocular nonce word using a letter "m" instead of an "r" and an "n") often involve mistaking a close-packed "cl" for "d," "LI" for a squarish "U," and "rn" for "m."

Word resolution is governed by the space between words, and creates a potential conflict that must be taken into account when measuring factors like reading speed. This stems from the fact that **justification** (the normal setting of type so that both left and right margins are vertically aligned, as in this book) is done by adjusting the space between words. Metal type was justified line by line, by a skilled eye, breaking long words at line ends with hyphens where appropriate. Some cruder digital systems like word processors still do this, and never revisit earlier lines when a later one proves hard to justify. More sophisticated systems like typesetters justify whole paragraphs at a time, roughly equalizing the space between words throughout the paragraph to start with, and then testing multiple hyphenation and spacing decisions to arrive at a smooth, even setting. This avoids **rivers** of white-space meandering up and down from line to line, distracting the eye while reading.

Finer details that can be measured include the type of font (**serif** or **sans serif**), the **type style** (Antiqua, transitional, classical, cursive, humanistic, etc.); line length; line spacing; x-height; and many other apparently innocuous but influential aspects of type design. Some of these are explained in more detail in the section on serif or sans serif fonts below.

In general, sticking to one or two typefaces per document is recommended unless you are going for the Victorian playbill effect (see Figure 5.5): too much shifting between fonts has been shown experimentally to affect perceptive efficiency (Sanocki, 1987).

Typographic design

Typographic design describes what the document should look like (also called the layout). Formal design specifications (a "compositor's spec") will explain every component element found in a document in technical terms to tell the compositor how to typeset it, or provide a template in software that can be used to automate the

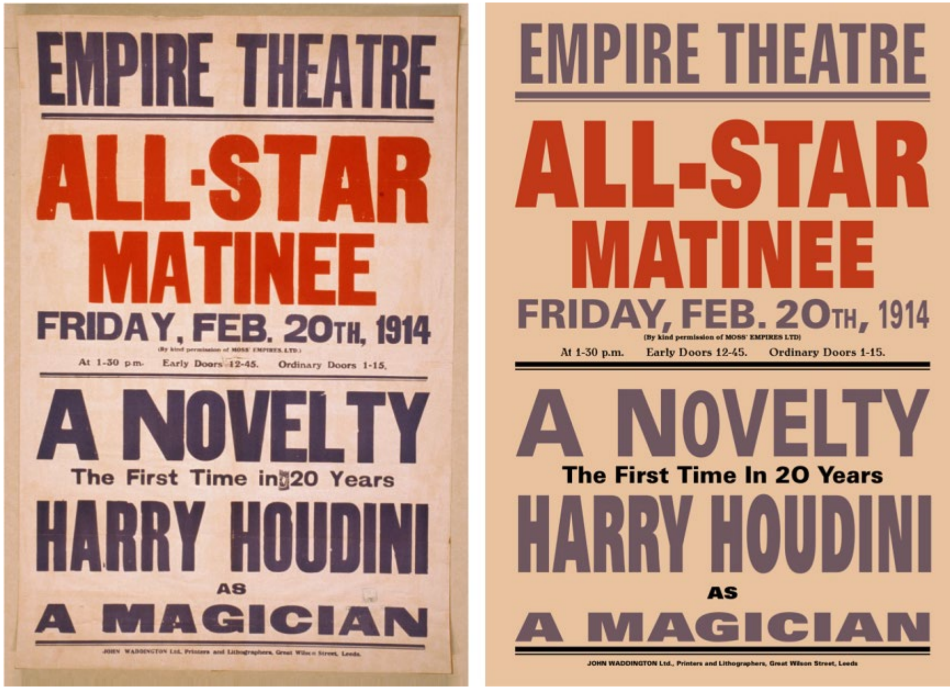


Figure 5.5 Early 20th century playbill (left) and draft digital typographic reconstruction (right) using modern fonts (Univers, Bookman, and Antykwa Torunska). *Source:* (left) from the copy in Library of Congress Prints and Photographs Online Catalogue, reproduction no. LC-USZC2-3781; (right): CC:BY-NC-SA © 2016 Author.

setting (see the section on consistency below) to ensure consistency. In more general terms, this covers (in addition to which fonts to use):

- the size and shape of the surface: office paper, paperback book, hardcover textbook, mobile web page, eBook reader, light display, etc., right up to the biggest 96-sheet billboard poster (about 10 m wide by 3 m high);
- the arrangement and spacing of the component elements (vertical or horizontal stress, sparse like a novel or complex like a textbook, the use of rules (lines), and the layout of headings, sidebars, lists, figures, tables, etc.);
- the "color" of the page (densely packed pages look darker than more spaciously arranged ones), as well as the use of actual color;
- the overall look and feel (elegant, attention grabbing, spartan, decorative, or whatever impression the designer wants to convey).

In particular, measurement of the usability of size and shape decisions in digital typography must take into account the growing need to repurpose published material—to recast it from a book into an eBook, or from a thesis into a Web site, or from a Web site into a printed brochure. Apart from the effect on spacing, discussed in the section on white space below, any changes to the size and shape will also affect the flow of text, and thus where lines end and what hyphenation is needed. Although most reflowing systems (notably on web pages and in eBooks, but also in PDFs) are reasonably effective, excessive change may damage legibility, such as trying to squeeze the text into columns that are too narrow. Tables in particular are relatively immutable compared with normal paragraphs: you can't squeeze and stretch them as easily, so any measurement of the usability of one format compared with another must allow for such limitations.

We mentioned at the beginning how, early in history, our conventional forms of layout started to appear, for common elements of a document like paragraphs, headings, lists, and tables. Over the centuries, readers educated in Western conventions have become accustomed to seeing these elements, so choosing well-known and instantly recognizable forms will contribute to the degree of perception.

For example, Western readers have been brought up to understand that when a line falls short of the right margin, and is followed by vertical white space or a line that is indented, it's a new paragraph, which signals the start of a separate unit of thought or argument. A line in bolder or larger type, with some vertical white space above and below, especially with a number prefixed, signals a heading or subheading: the start of a new section. Other visual cues are the indentation and numbers or bullets, which signal a list, and the visual separation of figures and tables from the surrounding text. Consistency in all these, which aids cognition, can be achieved by good composition.

Composition

This is the act of putting the characters together to create the textual content. Usually this is what the author writes, probably using a word processor. It may then be checked and corrected by an editor or proofreader (possibly using specialist software), and then put into its finished form by a compositor according to the designer's specification, which we mentioned in the previous section, using a typesetting system.

It is, of course, possible (and very common) for the whole document to be created, edited, and typeset by the author, but it usually shows. Authors are good at writing, and, if experienced or properly trained, they will be good at proofreading and editing, but very few are also typographers.

Originally, authors wrote longhand or on a typewriter, and the publisher arranged for everything else. Since the widespread use of digital systems, publishers try to insist on certain standards for what they will accept, such as Word, LaTeX, or HTML5 documents, and imply time penalties if authors send them CrudWriter Deluxe files that no one can read. These limitations are, in themselves, a measurable usability factor, as conversion into the publisher's typesetting system requires skill and a sensitivity to the context. In his advice to typesetters, Charles Fyffe wrote:

> It is true, of course, that the copy should be correct when it comes to you, but it rarely is; and few people have the head for detail of a good typographer. In general, if *you* don't, no-one else will correct the copy, and probably no-one else has the ability (Fyffe, 1969, p. 59).

Accuracy is, of course, essential, as is the knowledge of what can be done and what cannot. Measurements can therefore include compositors' decisions on formatting that were not or could not be foreseen (see Figure 5.6), and had to be made "live."

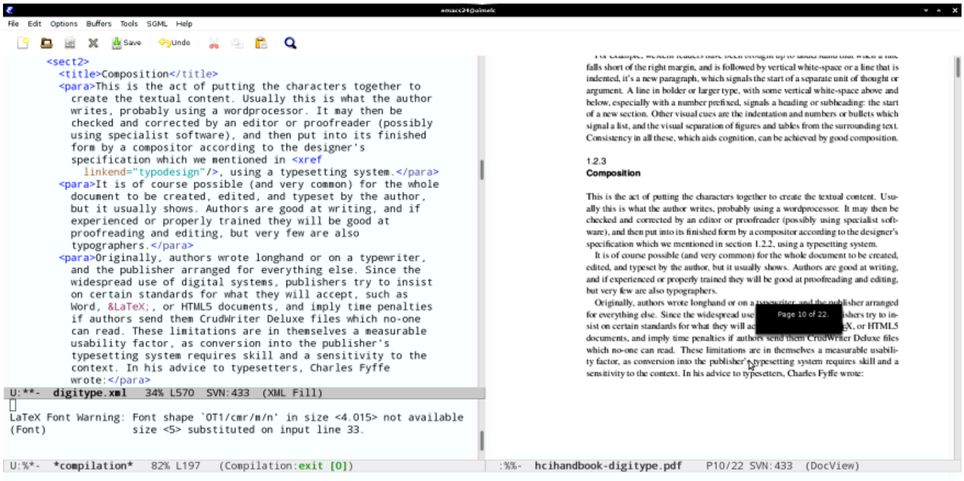


Figure 5.6 The act of digital composition: source text (left), WYSIWYG display (right), and log of the typesetting engine (lower left). *Source:* CC:BY-NC-SA © 2016 Author.

This includes "tweaks" or "touch ups" where last-minute adjustments to spacing or break points are made to prevent problems arising elsewhere (usually later) in the document. Some of the more useful of these are described in the section on tweaks and touch ups below.

Overall

The final outcome, the ultimate measure of success of the digital typography of a web or print design, is often nothing at all—that is, the reader should not even notice that the document has been "designed," let alone notice the individual design choices that have been made along the way.

The objective of almost all typography is invisibility: the document should convey the message with the minimum of interference. The moment readers start admiring the fonts, or the layout, or the choice of color or paper—or even hating them—then the battle has been lost: they are spending more time noticing the appearance than reading what the author has written.

There are a few major exceptions, of course: advertising and publicity material, where almost any trick is valid to grab the reader's attention, and a startling font or a crazy layout may be a positive advantage; and specialist publications where the typography is quite deliberately experimental—*WIRED* magazine is one very obvious example.

I'm sorry to say that one final exception is probably *you*—the members of the HCI design community: by the time you have studied and lived with fonts, typography, layout, and the minutiae of digital composition, you'll have joined the ranks of those who can't look at a document without proofreading it or commenting on its layout or choice of typeface. Notoriously, at the annual meetings of typographic software and similar organizations, significant amounts of time are spent proofreading the menu at dinner!

Guidelines

It is usually inappropriate to advise on matters of aesthetics where circumstances and the designer's brief may take precedence, and much of typographic design falls into this category. There are also circumstances where guidance based on best practice can be useful, and we can apply some of these criteria to the usability of digital type and typography. It is unwise to be too dogmatic about them but there are a few techniques you can use that appear to have stood the test of time (sometimes quite a short time, in the case of the Web).

Planning

In the same way that any project needs a definition, objectives, methodology, and a way to measure the results, a typographic designer (or in fact any user of digital type) will need a brief to work from, explaining who is going to use the document or Web site, what it is intended to achieve, and what technologies can be used, with any restrictions that exist.

However, most people probably don't have a typographic designer on call, and few people actually undertake projects that are exclusively typographic. Much more common are projects that have relatively small but important typographic components, such as periodic or final reports, a Web site, documentation, or publicity material, where you have to do the job yourself. You'll still need the same information, though, so it's a good idea to write it down before you start. For each project, I write it out in summary, on a single sheet of paper, and pin it to the workboard on the wall, where I can glance at it for reference any time. I find that planned documents or Web pages tend to be easier to manage than those allowed to grow in an uncontrolled manner.

Some of the techniques used in HCI can also be useful as ways of managing the digital typography of a project. For example, you may use **personas** (thumbnail personality sketches) of some idealized but representative and realistic users, and consider how they would react to the document type and layout as it progresses, in the same way as you would do in other usability aspects of a project. Ideally, of course, you would have a panel of real people, and use some of the standard test methodologies in the HCI field, but this is rarely practicable for digital typography except in very large-scale projects.

Vision and usability: What you see may not be what you get

Most of what is printed or published on the Web or in eBooks is put there to be seen. Braille, audiobooks, and other nontypographic channels have wonderfully opened up access to information for those with limited sensory perception, but the concentration in typography must inevitably be on the visual.

Sight is the form of perception that uses light. Visual acuity varies, as does the readability of Web sites and print documents, so there is no single measure of who can and who cannot read a given site; but the World Health Organization figures place the number of people worldwide with visual impairment at 285 million (World Health Organization, 2014), so it is useful to separate out some of the components of

6. Any reference (express or implied) in these Regulations to contractors, sub-contractors, construction operations or similar concepts created by the Act is to be read as including, in relation to times, circumstances or purposes in relation to which any corresponding concept which has been superseded by the Act had effect, a reference to the superseded concept.

Figure 5.7 Legible and readable? A clause (in grey type) from the online version of Schedule 1 to the UK's Income Tax (Construction Industry Scheme) Regulations 2005 at http://www.legislation.gov.uk/uksi/2005/2045/schedule/1/made. *Source:* Open Government Licence 2005 HM Government. Publishing (http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/)

typographic design where thoughtful application can make the difference between a readable document and an unreadable one (see Figure 5.7):

- Legibility: this determines how easily we can see the letters. Light intensity is important: for example, most signage needs to be well lit and of high contrast; most people sit in the light to read a print book or newspaper; but some types of display screen don't always work well in bright sunshine, as they emit light rather than reflecting it.
- Readability: this determines how easily we can make sense of what we see; that is, how we turn the letters our eyes perceive into words our brain can recognize. It is made easier if documents use well-defined letter shapes that conform to readers' mental models of "what letters should look like," and if the documents adhere to the commonly accepted practices for design and layout.

Fortunately, if people find a document hard to read, it is much easier to change fonts and layouts in a digital system than it was in metal type—the RNIB continues to have strong demand from visually impaired readers for large-type editions of popular books. Documents, whether print or web pages, need to be designed to work in the prevailing environment, for which you need data on your readers. You also need to know your readers' mental models of lettering, which may vary widely according to culture, age, education, literacy, social status or class, and possibly many other factors.

- Variety: there is normally a range of ways to present the different **elements** that make up a document. While variety applied consistently can help identify the type of information, such as a heading, variety for its own sake becomes novelty hunting: fun at first (which is why it is successful in ephemera such as advertising, like the poster in Figure 5.5), but rapidly becoming annoying or confusing for the reader.
- Contrast: this is the difference in the form of letters used to convey variety, including size, width, shape, and weight. Big and bold may mean a heading; italics may mean special significance (foreign words, emphasis, technical terms, and many other reasons); small capitals may mean personal names; whereas a smaller size may mean a comment aside, like a marginal note or footnote.

Human factors also contribute to the way digital typography is used to transfer information, and as we saw earlier, there are aesthetic factors that do not lend themselves to direct measurement (although their after effects may be measurable, such as whether people remember a product because its advertisements used an unusual typeface):

- Literacy: not strictly a typographic consideration, but worth bearing in mind. With a public audience whose level of literacy is unknown, keeping everything short and simple may help more people understand more easily. If you know that your audience has specialist knowledge or advanced literacy, you may be able to convey the information better using more technical terms or more complex constructions.
- Time constraints: busy people don't want to spend time on unnecessary frills, so the placement of key information first, separate from longer explanations, can help readers decide straight away if the document or Web page is relevant. By contrast, commercial imperatives often result in techniques to force the reader down a specific path that exposes them to other material first, before they can get to what they wanted.

In both these cases, it is not uncommon for decisions and corrections to be left to the person who does the design or typesetting, as we saw earlier. A designer or typesetter still has to be literate, as their customers may not be.

Serif or sans?

In the choice of typefaces, there have been countless A/B tests to see if **serif** or **sans-serif** fonts are easier (faster) to read, or lead to improved cognition but the jury is still out on this much-debated question.

The standard argument is that serif typefaces (ones with the little tailpieces to the uprights) are more legible for normal continuous text than sans-serif typefaces (ones without the tail-pieces). But some typographers disagree, arguing that the effect may be partly due to familiarity and expectation (Garfield, 2011, p. 60), a view shared by some usability practitioners (Poole, 2008).

In the absence of a formal guideline, convention is probably the better guide: serif fonts for bulk text, where the reader is going to be doing a lot of reading.

White space

The width of the line of type (known as the "set") will affect a number of other factors, including how much information your readers can take in at one glance. Some writers suggest 1½–2 alphabets' worth of characters (based on long experience) or about 12 words of average (nonspecialist) text; fewer in languages with longer compounds (like German), more in languages with more short words (like English and some Slavic languages) (Fyffe, 1969; Parikh, n.d.). In practice this is often conditioned by external requirements (pack size in the case of products; sheet size for printed documents; usable screen width in the case of web sites, and so on).

Line spacing (still known as "leading" from the practice of inserting strips of type metal between lines to space them further apart; and pronounced "ledding") depends on line width: the longer the lines, the wider the line space should be, to allow the reader's eye to be able to track from the end of one line to the start of the next. Many digital systems use a set of built-in factors as a default, such as TeX's multiplier of 1.2, but the actual value needs to be determined by experience or experimentation.

Line breaking in digital systems is a function carried out by the typesetting program, which usually optimizes spacing and minimizes the number of hyphenation breaks required. Occasionally, manual intervention is needed, especially in scientific text where very long compound words occur—and in "displayed" text: headings, titles, or slogans that have been centered. Word processors (and typesetters) don't understand that line breaks in displayed text need to be inserted at a natural break in the sense or meaning, not simply at the point where the margin would otherwise be overrun. An example I used a while ago (Flynn, 2012) was from my local convenience store, advertising:

HALF-PRICE
DESSERTS FROM
OUR IN-STORE
BAKERY

It is only ephemeral, but if there had been space, it would have been more readable, especially to passers by, if it had read:

HALF-PRICE DESSERTS
FROM OUR
IN-STORE BAKERY

Word spacing is a part of the process of justification, discussed earlier. Digital fonts usually contain parameters that affect the initial space, which is then varied by the justification, between certain limits. Too wide, and the text contains "holes," which cause the rivers of white space up and down the paragraph, referred to in the section on type design; too narrow, and the words risk running almost into each other.

Letter spacing, often known as "tracking," should be avoided. It has its place in the spacing of titles and headings, where the kern between letters may need optical adjustment, but has no place at all in the continuous text of a paragraph. Some cruder systems have an unpleasant habit of increasing letter spacing when justification has run up against the limits of word spacing, and the result simply makes the text disruptively harder to read.

Tweaks and touch ups

When the text of a document has been finished, and all the typesetting and layout done, there are usually some infelicities in the way headings and the first and last lines of paragraphs fall. Clearly, headings must be on the same page as the start of the first paragraph following them, otherwise reading is disrupted. But an isolated single line at the top or bottom of a page (respectively called a "widow" or an "orphan") is usually avoided for the same reason.

Again, software may try to avoid this by adjustment of the spacing on the previous or following page but the grid systems used in some page layouts may forbid this. Nevertheless, there are times when the stretching or shrinking of a page to accommodate just one line less or more is required in order to avoid an unwelcome break point. Typesetting software will normally break between pages as and when required but, in the above cases, a manual page break may be needed to force matters.

Color

Color used to be too expensive to use on arbitrary pages of printed documents but recent advances in reproduction technology have made it much cheaper. As Web pages have shown (where color is unaffected by the cost factor) it can be a very effective tool for highlighting specific elements of the document. But care needs to be taken to make sure that the tonal intensities are still valid for readers with varying forms of color blindness, otherwise the effect is lost on them.

The technique of using color for textual elements (known as "spot" color) in print documents still needs to be used with discretion, unless you are trying for a special effect (the example of *WIRED* magazine was used earlier). Color in the sense of color illustrations, known as "process" color, is a different requirement, and is not connected with the typographic use of color, except that the way in which books are made (folded from multiple sheets of paper) can influence whereabouts in the book color can be used economically.

A recent, beautifully crafted book petitioned for the use of full process color on all pages, arguing that the monks of mediaeval Western culture used it for the fabulous illuminated manuscripts of the era, and their equivalents in many other cultures (Kirschenbaum, 2005). The author believes it is time that we abandoned Gutenberg's monochrome typographic practices in favor of color and free-form type and layouts on every double-page spread of a book. It is a powerful and persuasive argument, but in the current economic circumstances still unaffordable. Its use on independently created and domestic documents, however, printed on ink-jet or laser printers in smaller quantities, is certainly a possibility.

In 2016 it was still very fashionable, especially among young or inexperienced designers, to create Web pages that use very small pale grey type on a white background. Readers with 20/20 vision may have no problem with this, but defects of vision increase with age, and if your readers are older or less visually acute, it may make the page or site almost illegible. Decisions on type size will also be influenced by this: 10 pt is conventional for books, 12 pt for business reports, and 9 pt is common for paperbacks to reduce the page count. If paper costs are not relevant (e.g. on the Web), larger type makes it easier to read.

Some familiarity with color theory is required to make consistent sense of a multi-platform digital document. On paper, color is a function of light reflecting from a surface; on screen, it is emitted light, shining into your eyes. The nature and quality of the light is different, and what may be apparently the same color when reduced to its Red–Green–Blue (RGB) code on a screen looks quite different from its Cyan–Magenta–Yellow–Black (CMYK) counterpart used in print.

Consistency

One of the keys to successful digital documents is consistency in the way that equivalent components are shown. This was also true in the nondigital era, but the task was the job of the craft compositor, not the individual author. Convention is a strong force, and convention has it that in a single document, all headings of the same type (chapter, section, etc.) must be done the same way (as they are in this book); all lists must be formatted with the same spacing and indentation; all figures must follow the same pattern; and so on. This is what gives a document what may be called its “professional” finish, in that an untrained author may be unaware that such conventions even exist.

In fact, consistency is remarkably easy to achieve but the technique is rarely taught at the time people learn to use their word processor. All such systems have a set of named “styles”—templates for appearance—and using them guarantees, for example, that all the elements you tag as Heading1 will come out formatted exactly the same. (It also ensures that converting the document to typesetter format or to a web page can also be done with almost 100% accuracy.) Most publishers have files of templates to do things in their style: I am using one as I write this, so whatever other faults there are in this chapter, inconsistency should not be one of them!

Contrast

Following on from the themes of color and consistency, headings on sections of a document are often set in a contrasting typeface to the normal text (for example, in a sans-serif type or in color). But when contrast is needed in midtext (for example, important words or phrases, or specialist terms such as those I have highlighted in **bold**), a radical change like color or size might disrupt the reader’s concentration. In this case italics is more conventional for some specific reasons, and bold for others, such as relative importance.

In academic texts, especially in the humanities, it is common to have maybe a dozen or more font styles in use every few words, within the flow of the text, and with specialist meanings for each: *italics* for editorial changes, **bold** for names, ***bold italics*** for keywords, SMALL CAPITALS for foreign words, and even underlining, a relic of typewriting not normally used in typesetting at all. The range of available typefaces is very large, and some have up to 30 font variants, not just bold and italic. Identifying this level of detail increases the utility of the text, as it conveys critical information to the expert without the need to signal it in any other way, but the usability then depends completely on the level of expertise of the reader.

Structure

Breaking up the information in a document into digestible chunks is generally regarded as a good way to convey your message (just as novelists break their stories into chapters). However, in highly structured documents (this book is an example), the hierarchy needs to be strict: chapters contain sections; sections contain subsections; subsections contain subsubsections; and so on. In this way, the headings themselves become affordances, increasing the usability and utility of the document, especially if they are hyperlinked from and to the table of contents.

You cannot break out of this with a chapter that dives straight into subsubsections without having previously passed through sections and subsections. Not only would it disrupt the numbering system, if one is used, but it would wreak havoc with the comprehensibility of the text. Without a template styling system that does the numbering, spacing, and font changes automatically, it is easy in a very long and complex document to lose track of what level you are writing at, which is why such documents need to go before an editor on their way to publication.

Margins

In the days of manuscripts, margins were used for notes (vellum and parchment were too expensive to waste). Print followed this pattern, and even now, all books except the cheapest of mass-market paperbacks tend to have up to 2–3 cm (an inch or so) of margin all the way round. Margins can appear asymmetric until you consider looking at a double-page spread all in one go: the extra "inner" margin is there to allow for the binding.

Margins should not be squeezed too tight—generous margins give a more relaxed appearance, and the reflective or contrasting nature of the extra paper or screen border appears to make it easier to read the page. There are all kinds of complex formulas for calculating margins, but a good starting point is about 5% of the screen width for Web pages or apps, and about 1 inch for U.S. Letter-size PDFs or 25 mm for A4 size.

Aesthetics

There has been a certain amount of research into the relationships between the aesthetics of Web pages (and to a lesser extent, print documents) and HCI—for a good introduction, see Tractinsky (2014); two interesting studies on the boundaries between interface and aesthetics and interface and usability are reported in Tractinsky, Katz, and Ikar (2000) and Tucha, Rotha, Hornbækb, Opwisa, and Bargas-Avilaa (2012); a rather different approach and conclusion are reported in De Angeli, Sutcliffe, and Hartmann (2006). However, studies tend unavoidably to involve users' *perceptions* of usability: as one report notes, "the lack of appropriate concepts and measures of aesthetics may severely constrain future research" (Lavie & Tractinsky, 2004).

Nevertheless, as these authors demonstrate, users' reactions to the aesthetics of a document or Web page are important, and the research identifies two dimensions that we can use to measure the effects: "aesthetic, pleasant, clean, clear and symmetrical" and "creative, using special effects, original, sophisticated and fascinating."

Another view has also been examined, again only in Web pages, finding a strong correlation between users' perception of visual complexity, structural elements (links, images, words, and sections) and "aesthetic appearance," which combines the two earlier dimensions (Michailidou, Harper, & Bechhofer, 2008).

But visual complexity may be a combination of the number of different structural elements and the overall design; this is why we need to consider variety as well as contrast.

Conclusion

When I started researching this chapter, I thought it might be a good idea to assemble all the research on the usability of digital typography first. There is, in fact, far more than could possibly be brought together in one place in the time available, but in any case, there is a strong line of continuity between the usability of traditional typographic practices and the digital practices—it's just that, in a digital system, changing things is so much easier and faster, and if you don't like them, they can be changed back just as easily.

Within limits, however. Systems like word processors are not (and never were) intended to replace the typesetter: they simply don't have the sophistication or automatability yet, although the gap is closing. In a word processor, if I want to change the body font (the one used for all the text) I either need to select all text (in which case everything else changes, including all my headings), or I have to use named styles, and modify each relevant one in turn. In a typesetter, I simply change the setting for the body font. The emphasis in word processors is to use the window / mouse metaphor for *everything*, regardless of whether it's appropriate, and this provides a powerful disincentive to using them for the picky fine detail of typesetting.

Someone asked recently on a mailing list if there was any measurement of how long people who used word processors spent adjusting the font and other drop downs (typeface, size, style, margins, etc.) compared with how long they spent actually *writing*. There turned out to be rather little on this, but one recent guesstimate from a respected contributor was, shockingly, 50%!

The power of digital typography is its capacity for automation, where appropriate. It's fun to play with fonts, but it's also good to get the work done efficiently and professionally.

Terms

- Character: a letter, digit, punctuation, logogram, emoji, or other symbol, as an abstract concept, unformatted; for example, a Euro sign (€), the capital last letter of the Latin alphabet (Z), the "hot beverage" emoji (☕) or "dog" in Chinese (狗).
- Codepoint: the official number allocated to every character in human discourse by the Unicode Consortium and defined in ISO 10646 (over 120,000 of them). Usually expressed in hexadecimal (0x prefix, base-16), for example, the € is 0x20AC, the Z is 0x005A, "hot beverage" is 0x2615, and "dog" is 0x72D7.
- Color: the density or darkness of the overall appearance of a full page of (assuming black type on a white background). When the density of material and the space around it looks balanced and even, the page is said to have "good color." Nothing to do with normal colors.
- Elements: the components or "pool" of blocks of text traditionally available to writers and designers. Among the most common are sections and subsections, headings, paragraphs, lists (containing items), tables (containing a caption, then rows made up of columns containing cells), figures (containing a caption and image), quotations, sidebars, panels, and footnotes. Consistency in the way the elements are presented increases the usability of the document.
- Font: the characters (or more accurately, glyphs) in a single style of lettering, like Times New Roman Bold Italic 14 pt. The term does not refer to a collection of styles: that's a typeface or a font family.
- Glyph: the actual physical appearance of a character in a particular font: what you see on the screen or on paper. Glyphs for the same character vary according to the font in use, for example the official euro design (€), "Z," a colored "hot beverage" emoji in Android apps, or "dog" in 6 pt Chinese (狗).

- Graphic designer: someone who designs mainly with images (pictures, drawings, photographs), but may also be a typographer as well.
- Leading: the vertical white space below a line of type to separate it optically from the next line. The wider the line, the more leading needed for the eye to return to the start of the next line comfortably.
- Sans serifs: typefaces whose letters do not have serifs, like Helvetica. No one seems to know the origin of these terms, except that there is obviously a French history there.
- Serifs: the little tailpieces on the ends of the upright strokes of letters in typefaces like Times. By extension, a serif typeface is one whose letters have serifs.
- Type designer: someone who designs typefaces.
- Typeface: a collection of related fonts belonging to the same family, usually available in roman (upright), italic (slanted), bold (heavier, blacker), and bold italic (a combination of the previous two).
- Type family: a collection of related typefaces varying in form but usually by the same designer. It may include dozens of variants like Light, Extra Light, Extra Bold, Extended, Expanded, Condensed, sans serif, Monospace, Decorative, etc.
- Type size / font size: a measure of the maximum height of the characters in a font, including the height of the stems (ascenders) on b, d, k, etc., the depth of the descenders on g, y, and j, and any other protrusions, like parentheses. Originally it was the height of the piece of metal on which the character was cast, which had to allow for the highest and deepest character in the font. Measured in points: 72.27 pt to the inch in traditional metal measurements, or 72 Adobe "big" points in most software (28.3464567 bp to the centimeter). Continental Europe uses a different size for traditional points.
- Typographic designer / typographer: someone who designs mainly with type: textual layouts, Web pages, books, articles, and other documents.
- UTF-8: the de facto standard for the way a codepoint is "encoded" in computer storage. Given that a single byte of storage can only represent a number between 0 and 255, most codepoints need several bytes to store them; for example, the euro sign is 0xE282AC, the letter Z is 0x5A, the "hot beverage" emoji is 0xE29895, and the Chinese for "dog" is 0xE78B97. Other encodings subsumed in UTF-8 include ASCII (the most basic of all) and ISO-8859-1 (more commonly known as Latin-1).

References

De Angeli, A., Sutcliffe, A., & Hartmann, J. (2006). Interaction, usability and aesthetics: What influences users' preferences? In *Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, DIS* (pp. 271–280). New York, NY: ACM.

Flynn, P. (2012). Typographers' inn. *TUGboat*, *33*(1), 8–10.

Fyffe, C. (1969). *Basic copyfitting*. London: Studio Vista.

Garfield, S. (2011). *Just my type*. London: Profile Books.

Kirschenbaum, V.(2005). *Goodbye Gutenberg*. New York, NY: Global Renaissance Society.

Lavie, T. & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-Computer Studies*, *60*, 269–298.

Lewis, J. (1963). *Typography: Basic principles.* London: Studio Books.

Loxley, S. (2005). *Type*. London: I. B. Tauris.

Michailidou, E., Harper, S., & Bechhofer, S. (2008). Visual complexity and aesthetic perception of web pages. In *Proceedings of the 26th Annual ACM International Conference on Design of Communication* (pp. 215–224). New York, NY: ACM.

Parikh, R. (n.d.). *Distribution of word lengths in various languages.* Retrieved from http://www.ravi.io/language-word-lengths

Poole, A. (2008). *Which are more legible, serif or sans serif typefaces?* Retrieved from http://alexpoole.info/blog/which-are-more-legible-serif-or-sans-serif-typefaces/

Sanocki, T. (1987). Visual knowledge underlying letter perception: Font-specific, schematic tuning. *Journal of Experimental Psychology*, *13*, 268–278.

Saunders, J. B, de C. M. and O'Malley, C. D. (1950). *The illustrations from the works of Andreas Vesalius of Brussels.* Cleveland, OH: World Publishing.

Tiersma, P. M. (2010). *Parchment, paper, pixels.* Chicago, IL: University of Chicago Press.

Tractinsky, N. (2014). Visual aesthetics. In M. Soegaard & R. Friis Dam (Eds.), *The encyclopedia of human-computer interaction*. Aarhus, Denmark: Interaction Design Foundation.

Tractinsky, N., Katz, A. S., & Ikar, D. (2000). What is beautiful is usable. *Interacting with Computers*, *13*, 127–145.

Tucha, A. N., Rotha, S. P., Hornbækb, K., Opwisa, K. & Bargas-Avilaa, J. A. (2012). Is beautiful really usable? *Computers in Human Behavior*, *28*, 1596–1607.

World Health Organization (2014). *Visual impairment and blindness.* Geneva: Author.

Part II

Design Process

6

Agile User-Centered Design

Gabriela Jurca, Theodore D. Hellmann, and Frank Maurer

Introduction

This chapter discusses the integration of agile software engineering (agile methods) and user-centered design (UCD). It presents a process model integrating techniques from both approaches and a set of general recommendations for performing agile UCD, and discusses potential pitfalls common in attempts to integrate agile methods and UCD. In our model for agile UCD, we emphasize the best aspects of both processes: an iterative, incremental approach to the design and development of a system that focuses on enhanced usability.

Agile methods

Agile methodologies are a collection of software development and project management techniques that emphasize a people-centered approach to the production of software. Agile methods emphasize frequent delivery of new and improved features to customers. This gives customers a chance to evaluate the implementation of new features on a regular basis and provide feedback. When new features do not meet customers' expectations, this feedback encourages the team to embrace changes that will create a product that better meets the customers' needs and is easier to use. The main measure of progress on an agile project is working software—if the customer does not like the way a feature behaves, it is deemed incomplete.

There are many different organizational and technical techniques used within agile projects but on a high level the key concepts are as follows:

- **Iterative and incremental development**: the project is broken down into small, discrete pieces—often called *user stories*—that can be completed entirely within a short time (in Scrum, a 1–4 week *sprint*). The idea is not to spend long periods of time to complete a feature entirely but rather to present a customer with the *simplest thing that could possibly work* or a *minimum viable feature*. The customer will be able to request changes, enhancements and extensions to features on a

The Wiley Handbook of Human Computer Interaction, Volume 1, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

regular basis, which means that the development team must plan to iterate over the design and code multiple times until a feature is complete. However, each iteration will provide a vertical slice through a system—from the user interface down to back-end components.
- **High-bandwidth communication:** face-to-face communication is preferred by agile teams because it gives all people involved in a project the opportunity to answer questions and resolve miscommunications quickly. A *customer representative* or *product owner* is appointed by the customer specifically to be available to answer questions from the development team and ensure that the product is aligned with the customer's needs.
- **Short feedback loops:** by meeting frequently with customers, and frequently delivering functional code for immediate use, issues with the product are discovered sooner. This is especially important for features that require major changes to the design or architecture of a system. Ideally, an agile project should strive to make use of *continuous delivery* and immediately release completed features to real customers throughout development. When things go wrong, *failing fast* helps towards speedy recovery.
- **Focus on quality:** on an agile project, the only measure of progress is working code, for the simple reason that nothing else has any direct value to customers. *Continuous integration, test-driven development, pair programming,* and other quality-assurance techniques are crucial to prevent nonworking code from being released to customers. The concept of *collective ownership* also ties into this: every team member is personally responsible for the success of the project as a whole, and is charged with improving the application wherever there are issues—even in code that was not written by them.

User-centered design

User-centered design (UCD) is an approach to designing highly usable systems by discovering and clearly understanding the needs of the user at the beginning of a project. It is more likely that a system will have high usability by understanding the system's users and designing interactions around their needs. From a development point of view, this approach is expected to reduce work and unnecessary features, resulting in a faster and less expensive development process. By creating a visual demonstration of what the finished system should look like, UCD also helps establish a shared *project vision* for what the product will be and how users will benefit from it.

As with agile methods, there are many techniques that are used within UCD but the key ones for our integration are:

- **User analysis**: the center of UCD is to understand the users. Designers accomplish this through workplace observations, interviews, focus groups, contextual inquiry, and similar techniques. In UCD, these activities result in the generation of a broader understanding of the system needed to be built because it includes aspects like other systems that users already employ, their level of technical expertise, and the way the new system will fit in to their existing workflow. As a result of user analysis, designers may create *personas*—descriptions of fictional users that represent a class of users of the system. These help create features that are designed with the users in mind.

- **Prototypes:** designers will also create *prototypes*—sketches or other images that represent what the user interface of a system will look like at a specific point in an interaction. Prototypes can be as simple as a sketch on the back of a napkin to get an idea of what elements need to be present in an interface, or as complicated as a set of user interface mockups that demonstrate each step in an interaction. Prototypes help demonstrate to developers what their implementation should look like, and can keep them focused on the implementation of only those features demonstrated, thus avoiding unnecessary work.
- **Usability evaluation:** because prototypes look somewhat like actual systems, it is possible to evaluate the usability of the system before the system even exists. In *Wizard of Oz Evaluation using low-fidelity prototypes*, the user is asked to interact with a paper prototype as though it is a working system. The user can indicate the elements of the interface with which they wish to interact by simply tapping on them, and the designer can show the user what the effect of their interaction would be by simply placing a new page of the prototype illustrating the result in front of the user. In this manner, designers can discover issues with the usability of the system before the system exists and make appropriate revisions.
- **Iterative design:** by iteratively improving the design of prototypes using feedback from usability evaluations, designers can come up with easy-to-use designs that will be more likely to meet customers' needs. Further, because the design of the system is refined before development, the interface design of the final system will be less likely to change drastically. Changes are made on prototypes, where it is cheap to revise the design, instead of in code.

Integration

The use of agile methods and UCD together makes intuitive sense: using the two techniques together could be expected to result in the frequent delivery and improvement of functionality with high usability. The design of the system would evolve through evaluations early on with real users providing feedback from actual use instead of artificial test scenarios. Related to this, the system's functionality would be more likely to meet its users' needs as users would be able to see and give input on designs for potential features earlier in the process. Perhaps most importantly, the project vision would be more cohesive—all people involved in the development process, including customers, would be able to see what the goal system would look like from a very early stage.

General Recommendations

In this section we present some practices that can be of use to teams using agile UCD.

Communication

Frequent and open communication between all stakeholders, the developers and the user-centered designers is very important for agile UCD integration, as most problems reported are related to a lack of communication (Kuusinen, Mikkonen, & Pakarinen, 2012). Some of the symptoms that may develop from lack of communication are time wasting, loss of important information, and developers that disagree with designs (Sy & Miller, 2008). Agile communication approaches, such as daily stand-up meetings (*scrum meetings*) and face-to-face communication, improve the efficiency of the team, as designs are better understood by developers and important information communicated between organizational units (Sy, 2007). Another technique to improve communication is colocation of all team members, so confusions and misunderstandings can be cleared up immediately.

Information radiators

An information radiator is a large, visible display of information about the current state of the development process (Meszaros & Aston, 2006). Information radiators should be placed in highly visible locations so that all team members can easily see them. The objective of an information radiator is to passively provide any observer quick and up-to-date information on the project, where crucial information is conveyed with just a glance. Examples of information radiators include scrum boards (which convey the state of tasks within an iteration at a glance), a whiteboard, or continuous integration dashboards and physical visualizations.

It is important that designs for the system be prominently displayed so that they can directly serve as information radiators. These designs should be updated as often as needed based on customer and user feedback. This will help keep everyone on the same page with respect to the design of the system and maintain a unified project vision. Another benefit is that information radiators increase communication and understanding within the team, allowing team members to confront problems early on. The designs can serve as a boundary object to help mediate communications between team members with different specialties or between different teams. At the same time, information radiators also allow for transparency with the customer.

Little design up front

Little design up front (LDUF) is an approach to UCD that provides only information that is immediately necessary for agile development to proceed (Adikari, McDonald, & Campbell, 2009). For teams using Kanban, this could be accomplished by performing design work as the first step towards each individual user story, while in Scrum this could be done by performing design work one iteration before the features being designed would be implemented.

The purpose of LDUF is to use a just-in-time approach to the contextual research and design work that occurs prior to the start of development of a feature. The benefit of LDUF is that it allows for trial and error, focus on just the next high-priority feature to be implemented, and reduced lead time between requirements elicitation and delivery of a feature to a customer. This compares favorably against traditional UCD approaches in which the design for the entire system is finalized before development starts. In the latter approach, there is a risk that it may then take a long time before customers use the application and notice these issues—making it costly to revise the design. In addition, when requirements are all determined at the beginning, some requested features might not be fully necessary by delivery time due to changes in the customer's understanding of the system or the work process itself.

Low-fidelity Prototyping

Sketches can be completed quickly—for example, a sketch of a possible design drawn on the back of a napkin with the customer at a lunch meeting. They should be simple, fast, and disposable. The key feature of a low fidelity prototype is how cheap it is to produce, which enables designers to rapidly experiment with a wide variety of different approaches to the design of a system (Dow et al., 2012). These prototypes should ideally be done with pen on paper to emphasize these characteristics (Buxton, 2007). The goal of low-fidelity prototypes is to encourage customers to provide honest and critical feedback on the interface without getting hung up on its appearance—color, font, round edges, and all the other styling issues that are easily changed during development. Customers and users are less likely to provide useful feedback if the design already looks finished. Additionally, because low-fidelity prototypes are intended to look rough and unfinished, they can be produced by anyone on the team.

Usability evaluation

While UCD emphasizes usability evaluations as a key way to increase the usability of a system, these evaluations are not an explicit part of an agile process. Evaluation of the usability of a system may occur before and after development (ideally, both). If usability evaluations are performed on a prototype before development, it will be cheaper to incorporate feedback as it is less costly to change a prototype than to change a developed and tested system. If usability testing is done before development, more effort can be spent on evaluating acceptance after development.

Two common types of usability evaluations are *inspections* and *empirical testing* (Madan & Dubey, 2012). Usability experts perform inspections, and there are many types of inspections that can be done. One popular inspection method is *heuristic evaluation* where experts provide their opinions and views on the system based on usability principles (Nielsen, 1994). Inspections can be time- and cost effective because users are not required for the process. The downside of inspection methods is that they are best performed by usability experts, need to be performed multiple times, and usually identify only minor problems (Jeffries & Desurvire, 1992). Inspection methods are best used early in the design or development process.

In empirical testing, naive users are presented with the system, and the responses of the system or the users are recorded in terms of different usability metrics. The benefits of empirical usability evaluations are that they tend to find problems that have a bigger impact on the users as compared to walkthrough methods (Karat, Campbell, & Fiegel, 1992). Empirical user evaluations are best performed at the end of a cycle, because they are more expensive to perform, yet highly valuable to determine the usability of the near-final system.

Wizard of Oz (WoZ) evaluations simulate a user's interaction with a system that is not fully implemented. A "wizard" controls the input and output of the system. Wizard of Oz evaluation can be implemented at different levels of fidelity (Liu & Khooshabeh, 2003).

One problem of WoZ evaluation is that it can be very resource consuming to prepare (Bernsen, Dybkjær, & Dybkjær, 1994). Also, WoZ may be cognitively exhausting on the "wizard," which may lead to validity issues, because the "wizard" might perform differently in each evaluation. Evaluations with paper prototypes may also be better suited for cases where the input is highly constrained, and where changes will not cause a large cascade of changes that are difficult to simulate (Davis, Saponas, Shilman, & James, 2007). However, WoZ provides many benefits, such as exploration of the interface and discoverability of flaws before implementation.

Some automated tools have been developed to decrease the effort required for WoZ, which also collect metrics from the evaluation, such as mouse movement and timing. For example, ASE facilitates distributed WoZ evaluations (Hosseini-Khayat, Hellmann, & Maurer, 2010), and SketchWizard facilitates the prototyping of pen-based user interfaces (Davis, Saponas, Shilman, & James, 2007).

Discount usability

Discount usability refers to methods popularized by Jakob Nielsen, which are designed to evaluate usability while minimizing cost and time (Nielsen, 1993, 1995). Discount usability methods are tailored to teams that have a smaller budget, in order to allow for some usability research as opposed to none. Some guidelines of discount usability include: no more than five participants per usability evaluation round (Nielsen, 2000), a focus on qualitative over quantitative data, paper prototyping, and heuristic evaluations. However, one problem that may occur when practicing discount usability is that experimental controls are not considered, which may lead to uncertainty of causation (Cockton & Woolrych, 2002).

Expectation management

When working with teams that may come from different backgrounds, it is important that everyone is aligned in terms of project expectations. For example, to stay true to agile concepts, everyone on the team should be aware that requirements, technology, and skill sets on the team may change, so the team needs to be able to respond to change. Expectations also need to be set with the customer, as to what level of involvement is necessary throughout design and development. It may also be important for the customer to understand that although changes to requirements are embraced in agile methods, each change comes with a cost—development work will still need to be done to make a change, and this work will take time, and it is not free. Agile methods also emphasize that developers should not be overloaded with work in order to meet deadlines, as it decreases morale, productivity, quality, and sustainability, so customers (and managers) need to be clear that developers cannot continuously work overtime to meet unrealistic deadlines.

Whole team

One crucial point for the entire development team to understand is that every person working on the project is individually responsible for the success of the entire project. In agile methods, this concept is known as "whole team." Each person should feel empowered to take action to improve any aspect of the project. If developers notice design issues in the system, for example, it is imperative for them to work with the designers to resolve those issues. Specifically with respect to code, this is known as the boy scout rule—always leave the codebase better than you found it (Martin, 2008). In agile UCD, it is understood that this rule applies to all aspects of the system.

At a higher level, this means that all team members are involved in all aspects of the project—especially planning. Every team member should be aware of how features are prioritized and what design decisions are being made. A key way of ensuring that all team members understand the status of the project is to have whole-team daily stand ups. In a stand-up meeting, all team members stand for a brief (maximum 15 minutes) status report where all team members answer three simple questions:

- What did I do yesterday?
- What will I do today?
- What problems are impeding my progress?

For design work, answering the first question should involve a quick presentation of new designs so that the entire team is aware of them.

Automated testing

The only measure of progress on an agile project is working code. This means that developers need to be confident at all times that the code they are releasing to customers is correct. It would be impractical to expect developers to manually test the entire system every time a part of it is changed, so automated testing plays a crucial role in software development.

From a developer's perspective, this means making sure that the system is built correctly—that the code in every method of every class in the system performs as expected. Unit tests can be written before the code that they test, a practice called test-driven development (TDD). In TDD, developers write a new test, run it to make sure it fails, and then write just enough code to get the new test to pass—known as doing the simplest thing that could possibly work. This allows developers to fearlessly refactor the codebase to iteratively improve their solutions over time: because they know that the test passed at one point, any subsequent changes that cause the test to start failing will be immediately apparent.

From a customer's perspective, it is often useful to create *acceptance tests*—tests that encapsulate a customer's expectations about how the system should perform at a feature level. Acceptance tests can also work with TDD to indicate when development of a feature is complete. Further, certain types of prototypes can be used in a TDD process to bridge the gap between design and completed development (Hellmann, Hosseini-Khayat, & Maurer, 2011).

Frequent deployment

Generally speaking, the more time elapses between when code is completed by developers and when changes are requested, the more expensive those changes will be. One way to reduce the time between when developers complete features and when customers are able to use those features and provide feedback, is to deploy features to customers very frequently. In Scrum, the ideal is to have deployable code ready at the end of each iteration, every 1–4 weeks, throughout the project.

In some situations, it is even possible to use automated deployment tools to deploy any commits to source control that make it through automated build and test processes.

This best case scenario is known as *continuous deployment* (Humble & Farley, 2010). However, continuous deployment is not suitable for all types of projects or customers, and relies on a very high level of discipline from all developers.

The benefits of this practice can extend to the usability of the system because the design of the product is evaluated as soon as possible. When customers have access to the system immediately, feedback can be collected for improving the usability of existing features.

Of course, it is important not to forget a financial benefit of early and frequent delivery. For a business-to-business development project, the release of functionality can be tied to the delivery of features. For a business-to-customer project, where software will be developed for users in the general public, it is possible to start charging for a project early on in order to fund further development (for example: the Steam Early Access system—see store.steampowered.com / earlyaccessfaq).

Iteration and incremental development

An agile UCD process is fundamentally both iterative and incremental. Small units of functionality—individual user stories—will be designed, implemented, and released to customers frequently. The team can also expect to iterate over the design and implementation of features multiple times in collaboration with customers to evolve the system towards a useful as well as usable solution.

Integrating Agile Methods and UCD

On a fundamental level, UCD seeks to perfect the design of a system's interface before development begins—an attempt to *prevent* change after a certain point in time. This is opposed to an agile approach, in which development should begin early in a project, and functionality and design are expected to evolve over time—an attempt to *embrace* change. The tension between these opposing goals can manifest during the development process.

To balance between these two perspectives, the compromise is to develop a concrete but high-level product vision at the start of the project.

High-level product vision

A vision is a high-level picture of the final outcome of the project. The vision captures what the customer expects from the product, and thus what should be achieved on an overall level. The vision provides a common definition of the project to everyone involved in the project. Everyone on the team—including stakeholders, developers, and designers—should be able to understand the vision clearly. A product vision is often delivered in the form of a vision statement, a product feature sheet, or a descriptive document.

It is important to have solid vision early on in the project, so that everyone can stay aligned to the true goal of the product as the project proceeds. The vision should be formed before the team members begin to work on the requirements of the product, so that the requirements can reflect more accurately the expectations of the customer.

The project vision can later be refined or revised when appropriate based on further insights gained during the project.

When there is a lack of vision, several dangers may occur. The first problem is that decision making will be slowed down. At every step of the process, the stakeholders, team leaders, developers, or designers may have conflicting ideas about what the product should entail.

Another problem that may occur with a lack of vision is scope creep. Without a clear understanding of the final product, the team will not know when to stop the inclusion of new features, or when to stop enhancing current features.

It is also important to note that designers are often not full-time members of development teams. In fact, designers often work on multiple teams at any one time, especially in smaller organizations. A solid product vision will help the designers to produce more timely and useful designs as they navigate between teams.

Multiple visions

Multiple visions only need to coexist for a project when that project has multiple distinct user bases. If the system has multiple user bases, then the members of the team need to understand what the vision for each user base should be.

In the case of multiple visions, more thought is required on how the visions should be developed. One of the issues that needs to be addressed is whether all of the visions need to be implemented in parallel, or if they should be implemented sequentially. Although the visions will be built differently, there might be an overlap between the systems, or a core system. In consideration of the budget, perhaps a core system may need to be developed first.

An integrated process model

The techniques described in the previous section are important to a successful integration of agile methods and UCD, but they can of course be implemented as multiple different process models. In this section, we explain what one such process model looks like. However, in a true agile process, the team may modify the development process to better suit its own needs.

The goal of the framework developed by Da Silva, 2012; Da Silva, Martin, Maurer, and Selbach Silveira, 2011; and Da Silva, Silveira, Maurer, and Hellmann, 2012, is to integrate UCD practices into an agile process, in order to deliver a highly usable system consistently. The high-level diagram of the framework is presented in Figure 6.1.

Product vision

The first step to take is to develop a clear product vision (Raison & Schmidt, 2013). As described previously, the vision helps to unify the goals for the product. The product vision then needs to be understood by all members of the team, including the stakeholders, the developers, and the UCD designers.

Iteration 0 Iteration zero is typically performed after the initial planning, but before development begins and includes initial design work, contextual inquiry as well as setting up development environments and tools for the project.

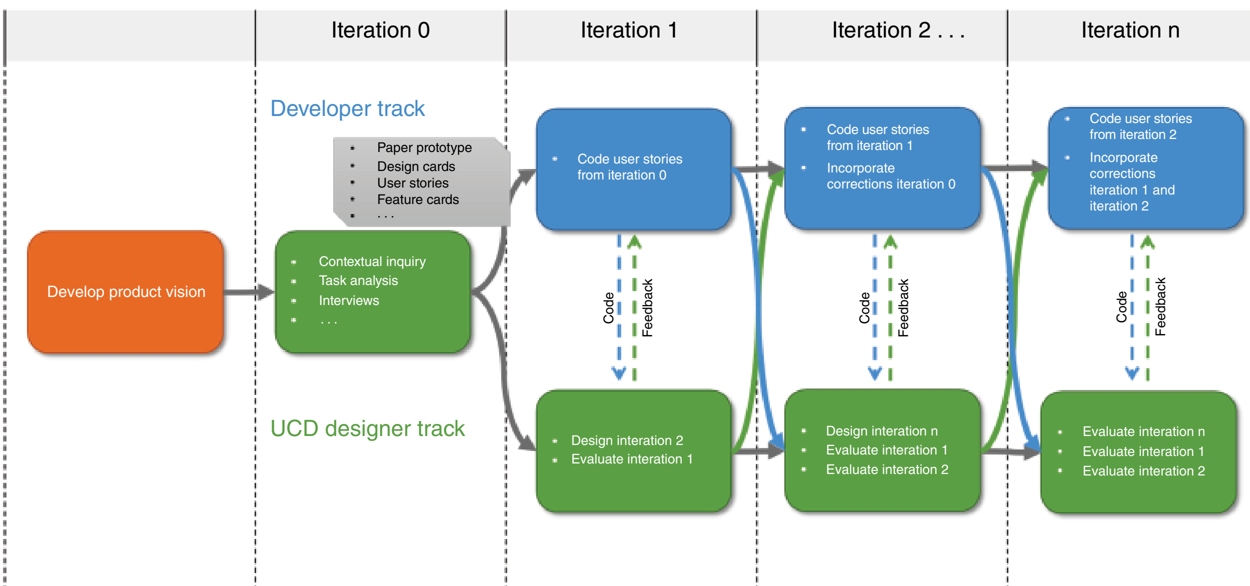


Figure 6.1 A high-level diagram of an Agile-UCD framework and associated artifacts (in the grey box).

Before development begins on the product, it is recommended to begin with a preliminary iteration, which allows UCD designers to conduct initial user research. User research is useful because it examines how a user interacts with a design or product, whereas developers tend to consider only what a customer wants. The problem with development based on what a customer wants, is that even the customer may often not know what he or she wants, or may not have the technical knowledge to know what to want. Therefore, the skills that user-centered designers have in gathering design requirements based on contextual research are the main asset of this iteration. User research is not finished after Iteration 0 and can be included when appropriate in later iterations.

Some of the practices for user research include observations, task analysis, interviews, personas, and iterative paper prototyping. During this stage, it is also helpful for the team to perform benchmarking of competitors' designs.

It is recommended that developers also participate in the contextual inquiry, since it may help the team members understand the "big picture" of the project. Other tasks that developers might perform during Iteration 0 may be related to development, such as setting up an informative workspace, databases, a continuous integration system, source control, and other technical set-up work. During this iteration, it is also important to set up the working environment to include information radiators, such as white boards or pin boards.

The idea of Iteration 0 is that it sets the UCD designers at least *one iteration ahead* of the development team, so that the UCD research and evaluation may be done before development. It is important to remember that we are adapting UCD to agile methods—which means that designs are done *just in time* and result in designs of the features in development instead of a design of the whole system. The benefit of "just-in-time" designs is that the designs will not be outdated by the time the development team can begin to work on them, as new designs will incorporate feedback from parts of the product that have already been developed and evaluated. That is, user stories should be gathered, specified, and designed for only the current and next iteration. The goal is to also prevent the team from being overloaded with story cards and designs for features that are not important until further in the future.

Artifacts produced by Iteration 0 may be paper prototypes, design cards, user stories, or feature cards. These artifacts are then used by developers in Iteration 1 in order to start development of the features.

Iteration 1 Developers may start to code the features by using the artifacts produced from Iteration 0 as well as their face-to-face interactions with designers on the team. It is important that the user-centered designers are collocated with the developers during this time, so that any misinterpretations or questions may be cleared up immediately. Corrections may also be implemented directly by consulting with the designers.

Meanwhile, the designers should help the developers understand the designs, as well as do research and evaluate designs for the next iteration. In this way, the designs produced during Iteration 1 can then be implemented during Iteration 2. Moreover, the designers should also begin to evaluate the usability of the features that have been implemented by the developers so far.

Design evaluation is a useful technique used in UCD, which allows designers to identify usability issues before effort is wasted on development. There is a cyclical relationship between design work and evaluation until the design is implemented by the developers.

Iteration 2 As in Iteration 1, the developers will begin to implement the designs made during the previous iteration. At the same time, any corrections that surfaced during evaluation of Iteration 1 should also be implemented.

The designers will continue to evaluate features implemented in Iteration 1 (preferably with real users, or through inspection), begin to design for the next iteration, and begin to evaluate the current iteration.

Further reading

The latest research shows that interest in agile-UCD integration has been increasing in the recent years (Jurca, Hellmann, & Maurer, 2014). Most of the research published on Agile-UCD also comes from experience and industry reports. Therefore, agile UCD is a hot topic for industry professionals, as the aspect of usability becomes more and more important in applications and technologies.

To meet with industry professionals and researchers who are currently trying to solve the problem of agile-UCD integration, the most popular conference venues are: Agile, CHI, XP, HCII, and DUXU.

References

Adikari, S., McDonald, C., & Campbell, J. (2009). Little design up-front: A design science approach to integrating usability into agile requirements engineering. In *Proceedings of the 13th International Conference, HCI International 2009* (pp. 549–558). San Diego, CA: Springer.

Bernsen, N., Dybkjær, H., & Dybkjær, L. (1994). Wizard of Oz prototyping: How and when. *CCI Working Papers in Cognitive Science and HCI*. Roskilde, Denmark.

Buxton, B. (2007). *Sketching user experiences*. Amsterdam: Morgan Kauffman.

Cockton, G., & Woolrych, A. (2002). Sale must end: Should discount methods be cleared off HCI's shelves? *Interactions, 9*(5), 13–18.

Da Silva, T. (2012). *A framework for integrating interaction design and agile methods*. (Unpublished PhD thesis). Pontifical Catholic University of Rio Grande do Sul, Porto Alegre. Retrieved from http://hdl.handle.net/10923/1487

Da Silva, T., Martin, A., Maurer, F., & Selbach Silveira, M. (2011). *User-centered design and agile methods: A systematic review*. Paper presented at the Agile Conference, Salt Lake City, UT.

Da Silva, T., Silveira, M. S., Maurer, F., & Hellmann, T. (2012). User experience design and agile development: From theory to practice. *Journal of Software Engineering and Applications, 5*(10), 743.

Davis, R., Saponas, T., Shilman, M., & James, L. A. (2007). SketchWizard: Wizard of Oz prototyping of pen-based user interfaces. *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology, Newport, RI*. New York, NY: ACM.

Dow, S., Fortuna, J., Schwartz, D., Altringer, B., Schwartz, D., & Klemmer, S. (2012). Prototyping dynamics: sharing multiple designs improves exploration, group rapport, and results. In H. Plattner, C. Meinel, & L. Leifer, *Design thinking research* (pp. 47–70). Heidelberg: Springer.

Hellmann, T. D., Hosseini-Khayat, A., & Maurer, F. (2011). *Test-driven development of graphical user interfaces: A pilot evaluation*. Paper presented at the International Conference on Agile Software Development (XP2011), Madrid, Spain.

Hosseini-Khayat, A., Hellmann, T. D., & Maurer, F. (2010). *Distributed and automated usability testing of low-fidelity prototypes*. Paper presented at the Agile Conference, Orlando, FL.

Humble, J., & Farley, D. (2010). *Continuous delivery: Reliable software releases through build, test, and deployment automation*. Upper Saddle River, NJ: Addison-Wesley.

Jeffries, R., & Desurvire, H. (1992). Usability testing vs. heuristic evaluation: Was there a contest? *ACM SIGCHI Bulletin, 24*(4), 39–41.

Jurca, G., Hellmann, T. D., & Maurer, F. (2014). *Integrating agile and user-centered design: A systematic mapping and review of evaluation and validation studies of Agile-UX*. Paper presented at the Agile Conference, Kissimmee, FL.

Karat, C.-M., Campbell, R., & Fiegel, T. (1992). Comparison of empirical testing and walkthrough methods in user interface evaluation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 397–404). New York, NY: ACM.

Kuusinen, K., Mikkonen, T., & Pakarinen, S. (2012). Agile user experience development in a large software organization: Good expertise but limited impact. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 94–111). Toulouse: Springer-Verlag.

Liu, L., & Khooshabeh, P. (2003). Paper or interactive? A study of prototyping techniques for ubiquitous computing environments. *CHI '03 Extended Abstracts on Human Factors in Computing Systems* (pp. 1030–1031). Fort Lauderdale, FL: ACM.

Madan, A., & Dubey, S. K. (2012). Usability evaluation methods: A literature review. *International Journal of Engineering Science and Technology, 4*(2), 590–599.

Martin, R. C. (2008). *Clean code: A handbook of Agile software craftmanship*. New York, NY: Prentice Hall.

Meszaros, G., & Aston, J. (2006). *Adding usability testing to an Agile project*. Paper presented at the Agile Conference, Minneapolis, MN.

Nielsen, J. (1993). *Usability engineering*. Cambridge, MA: Academic Press.

Nielsen, J. (1994). Usability inspection methods. *CHI '94 Conference companion on human factors in computing systems* (pp. 413–414). Boston, MA: ACM.

Nielsen, J. (1995). Applying Discount Usability Engineering. *IEEE Software, 12*(2), 98–100.

Nielsen, J. (2000, March 19). *Why you only need to test with five users*. Retrieved from http://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/

Raison, C., & Schmidt, S. (2013). *Keeping user centred design (UCD) alive and well in your organization: Taking an agile approach*. Paper presented at the International Conference on Design, User Experience, and Usability, Las Vegas.

Sy, D. (2007). Adapting usability investigations for agile user-centered design. *Journal of Usability Studies, 2*(3), 112–132.

Sy, D., & Miller, L. (2008). Optimizing agile user-centred design. *CHI '08 Extended Abstracts on Human Factors in Computing Systems* (pp. 3897–3900). Florence, Italy: ACM.

7

Ethnographic Approach to Design

Dave Randall and Mark Rouncefield

Introduction

And it is in understanding what ethnography is, or more exactly what doing ethnography is, that a start can be made toward grasping what anthropological analysis amounts to as a form of knowledge. This, it must immediately be said, is not a matter of methods. C. Geertz, *The Interpretation of Cultures*

Ethnography is not in any sense a unitary method but is a gloss on various and different *analytic* frameworks—thus there are Marxist, feminist, and postmodern ethnographies, all bringing different purposes to bear on inquiry. Here, we emphasize the role of ethnography for design-related matters. The issue of the relevance and purpose of ethnography is, we suggest, critical to understanding its role in human-computer interaction (HCI). In what follows, we provide a brief description of the history of ethnographic practice, a summary of the various theoretical and conceptual commitments that can accompany it, how it has been used in HCI/CSCW research, and finally some of the variants that have been argued for in the past few years as research has extended into new domains, of which the most obvious is the "virtual."

The Ethnographic Past

The origins of ethnography lie in a challenge to conventional theoretical thinking about other cultures, of which the best known example is probably Sir James Fraser's, *The Golden Bough*. Fraser's broad thesis, concerning the move from magic to religion and then to science, was untroubled by any direct empirical evidence. A succession of "classic" studies—Malinowski, Radcliffe Brown, Franz Boas—produced ethnographies of other cultures, which successfully challenged the established view of the "primitive mind." They were followed by a later generation working in the 1950s and 1960s, including Alfred Kroeber, Margaret Mead, Marshall Sahlins, and Karl Polyani.

The Wiley Handbook of Human Computer Interaction, Volume 1, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

Subsequent to that came the so-called "postmodern" turn exemplified by the likes of James Clifford and George Marcus, and (arguably) Clifford Geertz.

At the same time, theoretical developments that originally challenged the realist assumptions behind classic anthropological work have led to new perspectives on the relationship between the observer and the observed, and, in turn, have produced so-called "critical" ethnography; autoethnography, "reflexive" ethnography, institutional ethnography (Dorothy Smith), and so on.

Such a history is necessarily brief but serves one essential purpose. The history of ethnography in the social sciences is a history of changing convictions about the rationality of cultural members. Very roughly, early work of the structural- functionalist kind saw behaviors as serving purposes that ensured the stability of societies. This was gradually superseded by a "cultural" turn, which emphasized the need to see cultures largely in their own terms. The postmodern variant was a challenge to the very conception of a "culture," arguing that anthropology had presented an overhomogenized view and neglected the extent to which pictures of local culture were inflected by the anthropologist's own interests and concerns. We need not make too much of this here, save to point out that empirical investigation in HCI has traced a similar path. Early work saw the need for investigation as being concerned with requirements for organizational logic and efficiency, followed by the ethnographic turn, which consisted of an appeal to the details of work arrangements and that necessitated a more sophisticated view of the relationship between work and technology (inspired variously by Suchman and Wynn; the so-called Lancaster school, and the work of Heath and Luff), and latterly a more "reflexive" view, which has emphasized a more heterogeneous view of technology and work/life arrangements, and that has attempted to encompass notions of the "critical."

Ethnography and HCI

It is an error to see ethnography as a "method," as no particular steps are required for observation strategies to qualify. It is instead best thought of as a broad set of analytic commitments. These, roughly speaking, in the HCI context, are:

- Some kind of commitment to the naturalistic, meaning that we assume that the world is socially organized in ways that are specific to the settings in which people find themselves and set out to investigate that organization. It follows from this that ethnographers treat each setting as, at least to a degree, unique. As a caveat, "natural" does not refer to the ethnographer's presuppositions about what is natural or not, but to the way in which actors see the situation they are in. A laboratory is a natural setting for people who do laboratory work.
- An assumption that the setting and its activities make sense to its participants, and a commitment to understanding that "sense making" as best we can.
- An understanding of the setting as involving ordered, consistent sets of activity. Whatever is being done is typically done for good reasons and in familiar (to the participants) ways.
- Recognition of something akin to a "working division of labor"—although individuals perform activities, these are often embedded in interaction and cooperation with others. Understanding how cooperation and/or coordination is achieved is probably central to design related ethnography.

- A preference for detail over "gloss." This, of course, can mean many things, ranging from the close sequential work associated with conversation analysis to the general presumption that we can identify a typical order in which things are done based on an "egological" view of activity. This simply means that when people do things in some order, they are usually attentive to concerns like, "what should I do next?" "how do I best do that?" or "how should I go about this?" The point is that we work these things out in the context of (amongst other things) a set of responsibilities, relevancies, and rules whether we do so at work or elsewhere.
- An insistence on practical knowledge, constituted in skills, local knowledges, and competencies. It is held by ethnographers that there is a tendency to assume that tasks are mainly process driven, or routine, and hence entail little or no skill. Ethnographers have regularly demonstrated that the experienced and practised grasp of complex skills and knowledges is often "invisible" to casual observers.

Why Use Ethnography?

The ethnographic turn was founded on a series of, not to put too fine a point on it, complaints. These included that traditional HCI, while dealing more or less adequately with "usability" issues was failing to deal with the problem of "usefulness"; that traditional system design was spending too little time on the complex business of "requirements"; that new, distributed, technologies and new organizational arrangements required new methods of understanding, and that the dominant laboratory, or quantitative, methodologies lacked an ecological validity. A number of well publicized "disasters" have certainly contributed to this (the huge expense associated with the often unsatisfactory introduction of electronic patient records being one relatively recent example) suggestion that traditional methods of requirements elicitation were inadequate, or in need of supplementation, by methods better designed to bring out the socially organized character of work and other settings. Design is arguably a "satisficing activity"—more of an art than a science, which deals with messy indeterminate situations, and "wicked problems." If so, then before designers can solve a design problem they need to understand what they are designing for, what new technology might do, who might use it and how, and what circumstances might govern successful (or unsuccessful) introduction. Perhaps the main virtue of ethnography is its ability to make inroads into describing how "users" construe such issues. More specifically, the task is to make visible the "real world" sociality of a setting through detailed descriptions of the "workaday" activities of social actors within specific contexts and present a portrait of life as seen and understood by those who live and work within the domain concerned. It is the ability of ethnography to understand a social setting as perceived by its participants (for HCI, the archetypal users) that underpins its appeal. Its chief characteristic, supposedly, is the researcher's (prolonged) immersion in the setting, and the detailed observation of circumstances, practices, conversations, and activity that comprise its "real-world" character. Having said that, and as pointed out by Randall, Harper, & Rouncefield (2007), issues surrounding prolongation, detail, and so forth, are very much determined in practice by what it is that is being studied.

"Doing" Ethnography

There is a longstanding debate about the degree to which doing ethnography relies on specific or specialized skills (associated with anthropological or sociological training). It is our view that such debates often rather miss the point. Data collection, in and of itself, we believe is, in the main, trivially easy. Data is, more often than not, lying around in plain sight but no one has bothered to collect it. There is nothing special to look for, nothing to find that is hidden. The ethnographer's job is to listen to the talk, watch what happens, see what people do when, anywhere, to write it down, tape it, record what documents can be recorded, and so on. The sorts of things that can be collected and recorded include conversations, descriptions of activities, diagrams of places, job descriptions, memos, notices, graffiti, transcripts of meetings, war stories, and more. Hughes and Sharrock (2002, p. 6) suggest that:

> Another simple truth about ethnography is that, given access, you can very quickly collect far more data than you can ever possibly use: a day's work can generate several hours of audio or video tape recording. Nor is there really much meaning to the idea that some things are *crucial* data—ethnography is a pretty diffuse exercise with [characteristically] vague objectives, if indeed, they can be called objectives at all: often the aim is just to see and hear as much as you can, and to get as good a record of what you can see and hear as possible. In the ethnographic setting it is *all* data, though there is no sense to having *all* the data.

Data collection, then, has no esoteric character that requires immense amounts of training. Having said that, the experienced ethnographer will remind you that data comes in many forms, and can consist of conversations, descriptions of activities, diagrams of places, job descriptions, memos, notices, graffiti, transcripts of meetings, war stories, and more. Even so, much of ethnographic practice is simply about presenting oneself as a reasonable, courteous, and unthreatening human being who is interested in what people do and then shutting up, watching, and listening. Ethnography requires simple abilities, including an ability to listen, show an interest in what people do and what they have to say, and tolerate long periods of boredom. Ethnography is an immensely ordinary activity requiring ordinary, mundane skills. At the same time, as Button and King (1992) put it, "*hanging around is not the point.*" Ethnographic *analysis* does arguably require some degree of training, or at least a certain kind of sensibility. What kind of training that should be is a matter of some dispute.

What does an ethnographer do?

What an ethnographer does is what any other person in the organization being studied is likely to do—watching, talking, sitting in meetings, learning their way around the organization. And it is not difficult. Data is not hard to find.

In terms of how to behave, while a researcher cannot cope with every personal idiosyncrasy there are some commonsense principles of conduct for the ethnographer. These primarily involve recognizing that, for those in the setting, their commitment to what goes on there is their business, their job, their moral universe, and the fieldworker, no matter what his or her personal inclinations are, must respect this. The point of fieldwork is to understand how people organize their activities and the rationales they bring to them. People, in Garfinkel's words (1967) are not "judgemental dopes," meaning that when they act the way they do, they do so for reasons that make perfect sense to them in the setting they are in at that time. This, we would suggest, is the sine qua non of ethnographic work. Sometimes the ethnographer can bring commonsense to bear because the world they are investigating is already familiar to them (family life; behavior in public) but other times, "what is going on" will be less straightforward and more time consuming. The apparent "strangeness" can assist us in "rendering the familiar strange yet recognizable." Put simply, it forces one to pay attention.

Ethnographic analysis for HCI

Making sense of the materials collected is, of course, not a matter of making any sense or, worse, trying to find *the* sense of the materials as if they had only one sense. It is not that such materials have any intrinsic value; the material is valuable insofar as it can be *made relevant* or useful for what it can say about the social organization of activities. However, ethnographic research is directed toward some kind of research objective. Its purpose is to develop an understanding, an analysis of a setting that has some relevance to design choices. While the fieldworker needs to go into a setting with as few theoretical preconceptions as to what will be found there as is reasonable, this is a posture designed to further a research aim, in this case understanding particular aspects of everyday, routine life.

On the face of it, ethnographic description looks to be a fairly simple matter until one gets beyond the realist assumptions contained in classic views. Once one recognizes, however, as the Wittgensteinian scholar Anscombe (1979) did, that everything is "under a description"—in other words, it can be described in a potentially infinite number of ways—then the difficulties become clearer. Our problem is that of *relevant* description. These problems of relevance might include what the choice of setting is to be and how (and by whom) it is to be determined. The apparently simple question disguises some increasingly difficult choices in a world increasingly characterized by devices that support mobility, and, concomitantly, one in which much of our conduct takes place online. Are we to choose a setting as an exemplar of something, and if so, of what? What kinds of regularity, or indeed exceptional events, should we be interested in? Is the example to be driven by an interest in a particular kind of work or by a particular kind of technology, or by some combination of the two? How do we determine which behaviors and technology uses in the setting turn out to be interesting, and for whom? What level of detail might be required of us, and for what purpose? Again, these are not intended to be glib questions, for they go to the heart of many of the disputes about the role of ethnography in design. What is obvious is that, given ethnographic commitments, we simply cannot know in advance how to decide upon these matters. Neither are we likely, without a sophisticated knowledge of the domain, to be able to answer the kinds of "what if?" questions that designers are prone to ask. And that is rather the point—that a developing experience of the domain, its practices, and the way in which technologies are used, will begin to provide us with some answers.

Fairly early on in HCI and CSCW, the idea that ethnography might serve a range of purposes was recognized. The different uses of ethnography identified by Hughes, King, Rodden, and Anderson (1994) include:

- Reexamination of previous studies. Here previous studies are reexamined to inform initial thinking.
- "Quick and dirty" or "lightweight" ethnography. Here brief ethnographic studies are undertaken to provide a general but informed sense of the setting.
- Concurrent ethnography. This is the idea of an ongoing ethnography that adapts its focus over time. Here design is influence by an ongoing ethnographic study taking place at the same time as systems development.
- Evaluative ethnography. Here an ethnographic study is undertaken to verify, validate or evaluate a set of already-formulated design decisions.

These categories should not be read as if they were mutually exclusive ways of using ethnography; some of the uses could be, and were, harnessed together, and the differences between them seen as differences of emphasis rather sharp demarcations. Design is a matter of responding to contingencies of various kinds. Design objectives are various, and this will have a bearing on the role of ethnography. In other words, while not necessarily buying into the picture of the design process as a series of discrete, clearly delineated and phased steps, it undoubtedly has different objectives at different stages, and, accordingly, implications for how design needs to be informed by relevant information about the domain.

The value of ethnography in design is a matter of some controversy (cf. Anderson, 1994; Plowman, Rogers, & Ramage, 1995) because there are no panaceas for the problems of design, and arguably could not be. This would entail "design" having a universal character—which it self-evidently does not—and an entirely predictable problem-solution structure, which it equally evidently does not. We can only expect ethnography (or the sociology that may be associated with it) to have a modest utility to design, and the role of ethnography is primarily as an *informational input* into design, and, as such, only one source of information. The input can be of *critical* value insofar as it can advise the designer of actual practices of work and may clarify the role that actual practices play in the management of work, matters that may not normally be captured by other methods. In as much as a position on the role of ethnography in CSCW design has emerged it can be expressed in its ability to make visible the everyday nature of work. As Suchman writes "ethnographies provide both general frameworks and specific analyses of relations among work, technology and organization. Workplace ethnographies have identified new orientations for design: for example, the creation and use of shared artifacts and the structuring of communicative practices" (Suchman, 1995, p. 61).

It is the ability of ethnography to describe a social setting as it is perceived by those involved in the setting (the archetypal "users") that underpins its appeal to designers. In particular, it offers the opportunity to reveal the "practices" of users to which they may not themselves attend—because they take them so much for granted that they do not think about them—or which they cannot articulate because of the bureaucratic or power relationships within which they are placed. Ethnography is valuable in identifying the exceptions, contradictions, and contingencies of work (and other) activities, which are real conditions of the work's conduct but which will not (usually) figure in official or formal representations of that work. The concept of practice, then, has become dominant theme in ethnographic enquiry. Practices can be said to have a number of characteristics, independently of the domain in question, and we would suggest that they can be understood through a focus on the following:

- they are egologically determined;
- they are sequentially organized;
- they take place in a (working) division of labor;
- they are socially organized through “awareness” (or “heed”);
- they are constrained / afforded by material factors (including technology);
- they are ecologically (elegantly) organized;
- they are coordinated;
- they are skillful and / or knowledge based;
- they are constructed in rule-based routines but (sometimes) beset by “troubles”;
- they take place in a morally and politically circumscribed universe.

The assumption is that, although ethnographers are engaged in providing useful information, it is for *designers* to draw conclusions from the results. The kinds of changes to design that will result from this approach are intended to have an *incremental* rather than a *comprehensively transformative* effect. There is no *intrinsic* design significance to the results of an ethnographic study, for such significance must be *relative* to the nature of the design exercise itself, to the purposes, conceptions, methods, and plans of those making the design. Ethnography should, at least to begin with, be done independently of *design* preconceptions, distancing itself from the preoccupations, enthusiasms, and orientations of the designer, refraining from looking at the setting and its affairs “through designer’s eyes.” While there may be a tension between the designer’s and the fieldworker’s roles, this is a *positive* feature, something that is hardly likely to be destructive of good design, through highlighting the difference between good *abstract* design solutions, good *practical* design, and, ultimately, the social and political effects of design solutions. (Dourish, 2006). In this way, to paraphrase Max Weber (see e.g. Giddens, 2013), we may think of ethnography as being “design relevant” but not “design laden.”

Concepts and theories

Various perspectives have been brought to bear on the ethnographic project in the HCI context, of which the best known are probably grounded theory, activity theory, distributed cognition, interactional analysis, and ethnomethodological “studies of work.” We deal with them very briefly here only because they cast some light on the reasons for the ethnographic turn. They are theories only in a weak sense, and are better understood as conceptual frameworks aimed at “illuminating” understanding through the building of concepts. They are, nevertheless, all concerned (with the exception of interactional analysis and ethnomethodology) with the problem of generalization and all (again with the above exceptions) also concerned with matters of policy and design. Increasingly, the most commonly used (and misused) approach is that of grounded theory. It is as well to bear in mind a number of points about “grounded theory.” Firstly, it comes in a number of variants. Glaser and Strauss, who first formulated the notion, were part of a Chicago school tradition that stressed a preference for inductive over deductive analysis, and which had established a critique of statistical method in the social sciences (see Blumer, 1956). Glaser and Strauss provided an approach that was essentially *comparative,* and they used it to study a range of settings in medical work (Glaser and Strauss, 1965, 1968). Subsequently, as is well known, there was something of a split between the two, and different versions came into play. We need not concern ourelves overly with this, but should note that with the work of Strauss and Corbin (1998) there was a distinct turn towards the methodological foundations of grounded theory, and towards "coding" in particular. In passing, we can note that there are other, less fashionable (in HCI), approaches to generalization including so-called "metaethnography" (see Noblit & Hare, 1988).

In much the same way, participation has become a central theme in design-related enquiries. So-called participatory design (PD) introduces a kind of "standpointism" into HCI research, arguing that moral or political commitments are inevitable in research, and following on from this that the point of view of the user needs to be embraced (which, in some sense, is foundational to ethnographic research). Kensing and Blomberg (1998) discuss what the conditions might be for successful PD work. These include, in their view, a perspective on the politics of design; on participation; and on methods, tools, and techniques. Again, this is not the place to discuss PD in all its detail, but such an approach clearly establishes the idea that the views or needs of different stakeholders should be taken into account and that design-related research is as much a matter of understanding social arrangements as it is understanding technical feasibility. Indeed, one branch of PD is sometimes referred to as a sociotechnical systems approach. It is not too unrealistic to say that one of the things that characterizes the PD rhetoric is a "reflective" style, in which the relationship between researcher and subject is foregrounded, and used to account for "what was achieved" and "what was not." In this version, PD often refers to "mutual learning" as being constitutive of the design process (see for instance Bjerknes and Bratteteig, 1988a, b). In this respect, it is similar to some postmodern versions of ethnography.

Activity theory (see e.g. Kaptilinen and Nardi, 2012) and distributed cognition (see e.g. Hutchins, 1996) are two other approaches we might briefly mention. The former emphasizes the developmental character of cognition and emphasizes the role of tools in mediating cognition and the role of communities in learning. The latter, in some fairly similar ways, emphasizes the role of culture in structuring representation in contrast to orthodox cognitive psychology. Both purport to provide a systematic way of dealing with what would otherwise be (in their view) some rather vague concepts.

We can also mention interaction analysis (IA) and ethnomethodological studies of work. Interaction analysis (IA) has become very influential in areas such as user experience design. It is no single approach, but a broad commitment. It is associated with the work that came out of Xerox Parc, and with a number of well- known figures in the HCI and CSCW communities including Lucy Suchman, Brigitte Jordan, Jeanette Blomberg, and Charles Goodwin, as well as Christian Heath, Paul Luff, and others in Europe. It emphasized the value of the close video analysis of work sequences for design-related problems. Examples of this kind of work include Goodwin's study of an airport (formulating planes) and Heath and Luff's (1992) study of the London Underground. Jordan and Henderson (1994) point to a close relationship between IA and ethnography when they observe, "Our own practice has been to do videotaping in conjunction with ethnographic fieldwork". This is because, as Hester and Francis (2000, p. 208) put it:

> a concentration on matters of sequentiality and turn-taking alone whilst perfectly legitimate in themselves, cannot provide an adequate answer to the question of the recognisability of "institutional activities and identities."

A further, and very commonly cited, usage, is that of "ethnomethodological ethnography" (Dingwall, 1981) or "ethnomethodologically inspired ethnography" (Silverman, 1985). Ethnomethodology, it is important to note, does not imply a view of method. It refers to members' methods, not those of analysts. Dingwall, for example, outlines the following characteristics; accomplishing social order; specifying actors' models; suspending a moral stance; creating "anthropological strangeness," and depicting stocks of knowledge. Ethnomethodologically informed ethnography, then, is concerned with member's methods for accomplishing situations in and through the use of their local rationalities. For ethnomethodologically informed ethnographic enquiry, members and their orientations and experiences are central. A central precept of ethnomethodological ethnography is to aim to find the orderliness of ordinary activities, an orderliness accomplished by social actors, constructed with their common-sense knowledge of social order. Observation focuses on the places and circumstances where meanings and courses of action are constructed, maintained, used, and negotiated.

> We are concerned with how society gets put together; how it is getting done; how to do it; the social structures of everyday activities. I would say that we are doing studies of how persons, as parties to ordinary arrangements, use the features of the arrangement to make for members the visible organized activities happen. (Garfinkel, 1974, p. 16).

Regardless of the perspectival choices outlined above, there is a common view that the purpose of ethnography is to display something of the "real-world" character of activities. Ethnographic studies focus on "real world, real time" activity, following courses of action as they happen. This requires showing not just that *that* some setting is socially organized but showing in detail just *how* it is organized.

Questioning Ethnography

It is possible to argue that problems with ethnography have three distinct, though overlapping, elements. The first of them has to do with ordinary practical difficulties, most of which have been well documented (Randall, Hughes, & Shapiro, 1994), and generally to do with the problems of "getting in, staying in, getting out" as well as issues of access and "gatekeeping," reliability, validity, generalization, and so on.

The second has to do with disciplinary critique—anthropologists and sociologists can and do argue that ethnographic work can be both biased and partial. From within anthropology and sociology, associated with the reflexive turn, ethnography has been accused of privileging a White, Western, male "gaze" (Clough, 1992; Reinharz & Davidman, 1992), and of promoting a colonialist attitude (Said, 1978) telling us more about researchers, and their (usually his) attitudes, than the cultures they purport to describe. This kind of attack and charge has been repeated by various writers in HCI as well. From within the ethnographic "establishment," Hammersley (1990a) has argued that the tendency to treat ethnographic description as involving simple reproduction of the phenomena described is misleading and mythical. He stresses that such description is always selective. Consequently, and following the "reflexive turn," he suggests that the relevancies and values that structure any ethnographic description must be made explicit. This is paralleled in the postmodern critique of ethnography, which questions its claims to "neutral realism," arguing that, in writing ethnography, the researcher does not merely uncover or detail reality but creates it in the interpretive process of creating the text because "reality" does not exist to be discovered. While such arguments from social philosophy are not particularly germane here—those interested should see Clifford and Marcus (1986), Hammersley (1992), and Marcus and Fischer (1986)—it does raise the issue of purpose again. This "self-reflexive turn" implies that researchers should document their own actions, attitudes, and prejudices (their own stance in relation to design relevance), and consider how this might have affected the setting they investigate. This has a particular resonance when we examine the particularities of sensitive settings.

The third lies in critiques applied in the specific context of HCI, and have to do with the use of ethnography in design-related arenas. Historically, it has mainly been limited to small-scale, well defined and usually quite confined contexts, well suited to the observational techniques employed. Consequently, there are obvious issues to do with large-scale, highly distributed organizations. Similarly, in small-scale settings there tends to be a clear focus of attention for the participants, who are typically few in number, and in which there is a relatively clearly visible differentiation of tasks at one work site. Scaling such inquiries up to the organizational level or to processes distributed in time and space is a much more daunting prospect.

In a similar vein, ethnography has historically been a "prolonged activity" and although "quick-and-dirty" approaches have been developed, the timescales involved in ethnographic research are often unrealistic in a commercial setting where the pressure is typically for "results yesterday." Millen (2000) has addressed these issues in a well-known paper. Moving out of the research setting into a more commercial one also raises different sets of ethical responsibilities as well as making access to sites more vulnerable to the contingencies of the commercial and industrial world.

Perhaps the most significant of recent arguments about ethnography and design was prompted by Paul Dourish (2006) in his well-known paper, "Implications for design," a position that has been robustly criticized by Crabtree, Rodden, Tolmie, and Button (2009). For Dourish, the relationship between ethnography and design has been underexamined. There are two elements to this. Firstly, it has led to some naïve renderings of design implications towards the end of otherwise competent ethnographies but, and this is a slightly different argument, also to the naïve acceptance of what we will call a "service" relationship, which ignores the potential that ethnography has for a more critical—perhaps overtly political—role. There can be little doubt that the "implications for design" sections of published papers at, for instance, HCI can be naïve or at a high level of generality. It has led some to argue for ethnographic studies oriented to "not designing" (see, for instance, Baumer and Silberman, 2011). It is also true that the critique of ethnography has recently stressed the need to come to terms with wider issues—precisely the moral and political

(Becker, 1967, was arguably the first paper to examine this in detail). Rode (2011) has, for instance, written on the notion of "feminist HCI" whereas others have investigated the role of fieldwork for HCI in development (see, for example, Anokwa et al., 2009; Chetty and Grinter, 2007; Wyche et al., 2012). We do not propose here to adjudicate between those who assert a special role for anthropologists (see, for example, Forsythe, 1999) and those who deny it but would make the point that such disputes come down to arguments about purpose. Crabtree et al. argue for a rejection of analytic "privilege"—that ethnography cannot claim to provide accounts that are "superior" in virtue of the professional status of practitioners—and claim that we can only do solid, detailed, empirical work that others may not be minded to do for a variety of reasons.

The arguments concerning the appropriate way to conduct ethnographies, the relationship to design, and the need for "new" methods due to the rise of ubiquitous computing, have also featured, somewhat controversially, in Crabtree, Rodden, Tolmie, and Button (2009) and subsequently in *Deconstructing Ethnography* (Button, Crabtree, Rouncefield, & Tolmic, 2015). This work was seen as controversial because it critiqued supposedly new forms of ethnographic methods as abandoning any serious attempt to engage with the social as a precursor to design, and resorting to fashionable "literary" and rhetorical methods that merely produced "scenic design" recommendations. Whatever the merits of this position, it serves to illustrate further the continuing debate and argument within HCI over the nature and value of ethnography, and its relationship to design.

A further critique has been introduced by George Marcus, who has introduced the notion of the "multisited ethnography." Marcus is an anthropologist who has been at the forefront of thinking about the nature of ethnography, the way in which ethnographic materials are presented or conveyed, and the "usages" that ethnography can be put to for some time. He asserts the view that ethnography needs to be understood as *always being driven by particular analytic foci*, and argues further that the future lies in interdisciplinarity, which will provide new analytic tropes. The move to multisited ethnography, according to Marcus, is predicated on significant changes in the modern / postmodern world. The development of new information and communication technologies that provide very rapid information flow, the rise of the global "marketplace," the globalization of "culture," and the rise of new categories of homeless "nomads" in which new "structures of feeling," identities or sensibilities become prevalent, have all in some way problematized the single site. Multisited ethnography, it is argued, might provide a response to these changes / problems in a number of different ways, including prompting a new form of political and moral engagement, innovative methodological treatments, and a more sophisticated relationship between the construal of data and our understanding of the relevance of theory. He outlines a so-called "multisited" approach to ethnography which represents, he thinks, a return to comparative ethnography, but in a different way:

> comparison emerges from putting questions to an emergent object of study whose contours, sites and relationships are not known beforehand, but are themselves a contribution of making an account which has different, complexly connected real-world sites of investigation…In the form of juxtapositions of phenomena that have conventionally appeared to be "worlds apart" (Marcus, 1998, p. 86).

The most important feature of this argument is that the problems of interdisciplinary engagement are not problems of method. Multisitedness implies an eclectic approach to "method," and thus cannot (in any simplistic way) be about remedying the failure of other "methods." Nor are they problems of substantive discipline-specific concerns because the contemporary crisis in those concerns is precisely what leads him to interdisciplinarity. The "multisited" view of interdisciplinarity, then, leads us to reflect on problems of *empirical* relevance, of *conceptual* orientation, and of the role of *comparison*.

Randall, Harper, & Rouncefield (2005) have argued that these choices require "a particular open-mindedness about method, a thoughtful selection of concerns, and an artful refinement of disciplinary…sensibilities." Our point is tangential to this: that debates about these matters are considerably less important than professional interests would have us believe. When we tell a story about how people in some context organize their work activities, the information they rely on, and the things they are attentive to, we are not suggesting that we have dealt with all aspects of their world, that we have ordered them in importance, got everything right, couldn't have described things differently, or have been scrupulously "objective."

Special Cases and New Methods

The "postmodern" turn identified above had a further characteristic, which had to do with the problematizing of the "field", with its notion of easily identifiable spatial, geographical, and cultural boundaries. There has been an explosion of work into other areas since the early 2000s, including the domestic arena, public life, and online communities. Each, arguably, bring specific demands, which are reflected in issues such as choice of setting, ethical considerations, issues relating to privacy, specific problems of access, and so on.

Sensitive or Private Settings

Increased attention given to how we construe the "natives'" point of view, how difficult it might be to understand and describe ways of life that are substantially different, and who has a right to describe them, has particular resonance in the case of difficult, or sensitive, settings. The sensitivity of some areas has given rise to a fashion for "autoethnography." It is suggested, often quite rightly, that particular people's experience of research "on" them has often been less than happy. Conventional ethnographic methods, it is held, do not adequately capture the thoughts, feelings, and views of those they are researching—such as women, disabled people, ethnic minorities, and so on—thereby becoming one further aspect of disadvantage. (Dartington, Miller, & Gwynne, 1981; Miller & Gwynne, 1972). Hence, "Disabled people have come to see research as a violation of their experience, as irrelevant to their needs and as failing to improve their material circumstances and quality of life" (Oliver, 1992, p. 105).

Autoethnography is proposed as one solution. It has generated an enormous amount of comment, both approving and otherwise. Wrapped up in it are postmodern concerns with reflexivity, political objectives and objectification. Hence: "Autoethnography is… research, writing and method that connect the autobiographical and personal to the cultural and social. This form usually features concrete action, emotion, embodiment, self-consciousness, and introspection" (Ellis, 2004, p. xix). "Autoethnographic texts… democratize the representational sphere of culture by locating the particular experiences of individuals in tension with dominant expressions of discursive power" (Neumann, 1996, p. 189).

It thus might entail personal narrative and experience; poetry; novelistic accounts and politics. It addresses some obvious themes; that ethnography is never wholly "innocent"; it can be used for "standpoint" purposes (and has been, most notably in the context of disability studies), and it recognizes the essential reflexivity between ethnographer and his/her subject. But then, as Atkinson (2006, p. 401) points out:

> The list of ethnographic projects that draw on a personal commitment or accident is a long one and does need to be extended ad nauseam. There is, therefore, no need to rely exclusively on postmodernist rationales to justify such auto/biographical bases for ethnographic work. The ethnographer's identity and the subject matter of her or his chosen research site(s) have long been implicated in one another, and it is not a new development in the field of the social sciences.

Other strategies have not sought a new analytic orientation but have, instead, adopted different methods, largely in order to deal with domestic and similar contexts. These include the use of "cultural" and "technology" probes, and the use of so-called "living labs." Probes are broadly equivalent to introducing artefacts of one kind or another into the field. They constitute a nonintrusive way of collecting information that might otherwise be difficult to obtain (see e.g. Crabtree et al., 2004; Boehner et al., 2007). They may also constitute a way of introducing a more creative or playful element into the field (see e.g. Gaver et al., 2004). They may be as simple as diaries, postcards or recording devices, or may entail wholly new artefacts. "Living labs" are not especially new, being associated originally with von Hippel (1976) and product design. They have been fairly widely adopted, however (see, for example, Abowd et al., 2000) and explicitly entail the introduction of new technology into settings like the home, or an equivalent, and are often associated with ethnographic analysis (see, for instance, Ley et al., 2014).

Mobilities

Another extension of the ethnographic project in HCI has been into behaviors with mobile technology, and the use of public spaces. The advent of so-called "smart phones" has extended the possibility for enquiry even more (see Brown, Green, & Harper, 2002; Büscher, Urry, & Witchger, 2010) and it can be argued that existing methods have dealt rather poorly with transitory phenomena—the movement of people, of objects, of information, and so on. The degree to which this might impact on the kinds of strategy that ethnographers employ in mobile contexts remains, as yet, unclear, although Buscher has suggested that the following consequences for method might become salient: observation of people's movements, accompanying people, the use of mobile video, diaries that record movements (including time-lapse cameras), observation of texting, discussion groups, and even older artifacts such as postcards, the creative use of location-based technologies, position tracking, capturing "atmosphere," engaging with memories, and examining behavior "en route."

The Virtual

Of course, the main change associated with our social life has been the unprecedented development of our use of the Web, and most notably the kinds of social media associated with Web.2.0 developments. This has given rise to a range of ethnographic practices including, for brief mention, digital ethnography (Dicks, Mason, Coffey, & Atkinson, 2005); cyberethnography (e.g. Miller and Slater, 2001); "virtual" ethnography (Hine, 2000); netnography (Kozinets, 2006), and so on. The latter is slightly different from the others in that it is the application of ethnography to online settings as a vehicle specifically for consumer research. Roughly speaking, there are two different issues to contend with here. One has to do with the fact that we might be studying the online behavior of others, while the second has to do with the use of digital tools of one kind or another by researchers. Digital ethnography primarily deals with the second of these two, while cyberethnography can be said to deal with both. In studying online behavior there are some self-evidently different features to contend with, notably that behavior is no longer colocated, face-to-face, or sequentially organized in quite the same way (a function of the fact that most, though by no means all, online behavior involves text rather than speech). There are undoubtedly some practical issues around how to study online behavior (these are discussed inter alia by Hine, 2000; Geiger & Ribes, 2011) but whether they are different *in kind* is a wholly different matter. There are differences of opinion about the analytic uniqueness of the online world, but we tend towards the less dramatic view. Ethnographers, after all, have always had to contend with communication at a distance; with interrupted observation, with textual or documentary analysis, and so forth. The problems become unarguably more pronounced in certain circumstances but they remain the problems of understanding interactional processes. Having said that, there are some features of online interaction that do seem to require the adoption of a more "mixed" methodology. Thus, for instance, the existence of "log" information affords possibilities that did not exist before, especially for comparing subjective views with objective facts. In many respects, differences of opinion reflect the disciplinary issues we mention above. For anthropologists, conceptual work may involve, for instance, reflections on the nature of the "self," "identity," or the role of "community," all standard social scientific terms, the meaning of which may alter in the online universe (see, for example, Wellman et al, 2001). The same might be said of concepts such as social network and social capital, commonly used by HCI researchers analyzing, for instance, Facebook behaviors (e.g. Ellison, Steinfield, & Lampe, 2007). Whether the use of such concepts matters in relation to design is a different matter. There are also some ethical issues that look rather different when doing online research as well. The doctrine of informed consent is difficult (to say the least) to apply, and there are issues around anonymity and privacy.

In sum, then, ethnography has undergone a number of transformations as a result of the focus on technology that characterizes HCI and similar areas. Although these transformations can be seen both in the deployment of relatively new methods and in the service of new theoretical orientations, we have argued that they do not alter the fundamental *analytic commitments* that underpin ethnographic work.

References

Abowd, G., Atkeson, C., Bobick, A. F., Essa, I. A., MacIntyre, B., Mynatt, E. D., & Starner, T. E. (2000, April). *Living laboratories: The future computing environments group at the Georgia Institute of Technology*. Paper presented at the Conference on Human Factors in Computing Systems (CHI 2000), The Hague, Netherlands.

Anderson, R. J. (1994). Representations and requirements: The value of ethnography in system design. *Human-Computer Interaction, 9*, 151–182.

Anokwa, Y., Smyth, T. N., Ramachandran, D., Sherwani, J., Schwartzman, Y., Luk, R.,… DeRenzi, B. (2009). Stories from the field: Reflections on HCI4D experiences. *Information Technologies and International Development, 5*(4), 101.

Anscombe, G. E. M. (1979). Under a description. *Noûs, 13*(2), 219–233.

Atkinson, P. (2006). Rescuing autoethnography. *Contemporary Ethnography, 35*, 400.

Baumer, E. P. S., & Silberman, M. (2011). When the implication is not to design (technology). In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2271–2274). New York, NY: ACM.

Becker, H. (1967). Whose side are we on? *Social Problems, 14*, 239–247.

Bjerknes, G., & Bratteteig, T. (1988a). The memoirs of two survivors: or the evaluation of a computer system for cooperative work. In *Proceedings of the 1988 ACM Conference on Computer-supported Cooperative Work* (pp. 167–177). New York, NY: ACM.

Bjerknes, G., & Bratteteig, T. (1988b). Computers—utensils or epaulets? The application perspective revisited. *AI and Society, 2*(3), 258–266.

Blumer, H. (1956). Sociological analysis and the "variable." *American Sociological Review, 21*(6), 683–690.

Boehner, K., Vertesi, J., Sengers, P., & Dourish, P. (2007). How HCI interprets the probes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1077–1086). New York, NY: ACM.

Brown, B., Green, N., & Harper, R. (2002). Wireless world: Social and interactional aspects of the mobile age. Berlin: Springer.

Büscher, M., Urry, J., & Witchger, K. (Eds.). (2010). *Mobile methods*. London: Routledge.

Button, G., & King, V. (1992). *Hanging around is not the point*. Paper presented at the workshop on "Ethnography in Design," *CSCW '92*, Toronto.

Chetty, M., & Grinter, R. (2007). HCI4D: HCI challenges in the global south. In *CHI'07 Extended Abstracts on Human Factors in Computing Systems* (pp. 2327–2332). New York, NY: ACM.

Clifford, J., & Marcus, G. (Eds.). (1986). *Writing culture: the poetics and politics of ethnography*. Berkeley, CA: University of California Press.

Clough, P. T. (1992). *The end (s) of ethnography: From realism to social criticism*. New York, NY: Peter Lang.

Crabtree, A., & Rodden, T. (2004). Domestic routines and design for the home. *Computer Supported Cooperative Work, 13*(2), 191–220.

Crabtree, A., Rodden, T., Tolmie, P., & Button, G. (2009). Ethnography considered harmful. In *Proceedings of CHI' 09 Conference on Human Factors in Computing Systems* (pp. 879–888). New York, NY: ACM.

Dartington, T., Miller, E. J., & Gwynne, G. V. (1981). *A life together: The distribution of attitudes around the disabled*. London: Tavistock.

Dicks, B., Mason, B., Coffey, A., & Atkinson, P. (2005). *Qualitative research and hypermedia*. London: Sage.

Dingwall, R. (1981). The ethnomethodological movement. In G. Payne, R. Dingwall, J. Payne, & M. Carter (Eds.), *Sociology and social research* (pp. 124–138). London: Croom Helm.

Dourish, P. (2006). Implications for design. *Proceedings of CHI '06 Conference on Human Factors in Computing Systems, Montreal, Canada* (pp. 541–550). New York, NY: ACM.
Ellis, C. (2004). *The ethnographic I: A methodological novel about teaching and doing autoethnography*. Walnut Creek, CA: AltaMira.
Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook "friends": Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*, *12*(4), 1143–1168.
Forsythe, D. E. (1999). "It's just a matter of common sense": Ethnography as invisible work. *Computer Supported Cooperative Work (CSCW)*, *8*(1–2), 127–145.
Garfinkel, H. (1967). *Studies in ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall.
Garfinkel, H. (1974) On the origins of the term "Ethnomethodology." In R. Turner (Ed.), *Ethnomethodology* (pp. 15–18). Harmondsworth: Penguin.
Gaver, W. W., Boucher, A., Pennington, S., & Walker, B. (2004). Cultural probes and the value of uncertainty. *Interactions*, *11*(5), 53–56.
Geiger, S., & Ribes, D. (2011). Trace ethnography: Following coordination through documentary practices. In *Proceedings of the 2011 44th Hawaii International Conference on System Sciences* (pp. 1–10). Washington, DC: IEEE Computer Society.
Giddens, A. (2013). *Politics and sociology in the thought of Max Weber*. Oxford: John Wiley & Sons, Ltd.
Glaser, B. G., & Strauss, A. L. (1965). *Awareness of dying*. Piscataway, NJ: Transaction Publishers
Glaser, B. G., & Strauss, A. L. (1968). *Time for dying*. Chicago, IL: Aldine.
Hammersley, M. (1990a). *The dilemma of qualitative method: Herbert Blumer and the Chicago tradition*. London: Routledge.
Hammersley, M. (1990b). What's wrong with ethnography? The myth of theoretical description. *Sociology, 24*, 597–615.
Hammersley, M. (1992). On feminist methodology. *Sociology, 26*(2), 187–206.
Heath, C., & Luff, P. (1992). Collaboration and control: Crisis management and multimedia technology in London Underground control rooms, *CSCW: An International Journal, 1*(1–2), 69–94.
Hester, S., & Francis, D. (2000). Ethnomethodology, conversation analysis, and "institutional talk." *Text-Interdisciplinary Journal for the Study of Discourse*, *20*(3), 391–413.
Hine, C. (2000). *Virtual ethnography*. London: Sage.
Hughes, J., King, V., Rodden, T., & Anderson, H. (1994). Moving out of the control room: Ethnography in system design. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work, Chapel Hill, North Carolina* (pp. 429–438). New York, NY: ACM.
Hughes, J., & Sharrock, W. W. (2002). Ethnography in the workplace: Remarks on its theoretical bases. *teamEthno Online, 1*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.463.6724&rep=rep1&type=pdf
Hutchins, E. (1996). *Cognition in the wild*. Cambridge, MA: MIT Press.
Jordan, B., & Henderson, A. (1995). Interaction analysis: Foundations and practice. *The Journal of the Learning Sciences 4*(1), 39–103.
Kaptelinin, V., & Nardi, B. (2012). Activity theory in HCI: Fundamentals and reflections. *Synthesis Lectures Human-Centered Informatics, 5*(1), 1–105.
Kensing, F., & Blomberg, J. (1998). Participatory design: Issues and concerns. *Computer Supported Cooperative Work, 7*(3–4), 167–185.
Kozinets, R. V. (2006). Netnography 2.0. In R. W. Belk (Ed.), *Handbook of qualitative research methods in marketing* (pp. 129–142). Cheltenham: Edward Elgar Publishing.
Ley, B., Ogonowski, C., Mu, M., Hess, J., Race, N., Randall, D., … Wulf, V. (2014). At home with users: A comparative view of Living Labs. *Interacting with Computers, 27*(1), 21–35.
Marcus, G. & Fischer, M. (1986). *Anthropology as cultural critique*, Chicago, IL: University of Chicago Press.

Millen, D. (2000). Rapid ethnography: time deepening strategies for HCI field research. In *Proceedings of the Third Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques* (pp. 280–286). New York, NY: ACM.

Miller, D., & Slater, D. (2001). *The Internet: An ethnographic approach.* Oxford: Berg.

Neumann, M. (1996). Collecting ourselves at the end of the century. In C. Ellis & A. Bochner (Eds.), *Composing ethnography: Alternative forms of qualitative writing* (pp. 172–198). Walnut Creek, CA: AltaMira Press.

Miller, E. J., & Gwynne, G. V. (1972). *A life apart: Pilot study of residential institutions for the physically handicapped and the young chronic sick.* London: Tavistock.

Noblit, G. W., & Hare, R. D. (1988). *Meta-ethnography: Synthesizing qualitative studies.* London: Sage.

Oliver, M. (1992) Changing the social relations of research production? *Disability, Handicap and Society, 7*(2), 101–114. Plowman, L., Rogers, Y., & Ramage, M. (1995). What are workplace studies for? In H. Marmolin, Y. Sundblad, & K. Schmidt (Eds.), *Proceedings of the Fourth European Conference on Computer-Supported Cooperative Work* (pp. 309–324). Stockholm, Sweden: Kluwer Academic Publishers.

Randall, D., Harper, R., & Rouncefield, M. (2005). Fieldwork and ethnography: A perspective from CSCW. In *Proceedings of the First International Ethnographic Praxis in Industry and Commerce (Epic) Conference* (pp. 81–99). Redmond, WA: Interactive Design Foundation.

Randall, D., Harper, R., & Rouncefield, M. (2007). *Fieldwork for design.* Berlin: Springer.

Randall, D., Hughes, J., & Shapiro, D. (1994). Steps toward a partnership: Ethnography and system design. In J. Goguen and M. Jirotka. *Requirements engineering.* San Diego, CA: Academic Press.

Reinhartz, S., & Davidman, L. (1992). *Feminist methods in social research.* New York, NY: Oxford University Press.

Rode, J. A. (2011). A theoretical agenda for feminist HCI. *Interacting with Computers, 23*(5), 393–400.

Said, E. (1978). *Orientalism.* New York, NY: Vintage.

Silverman, D. (1985). *Qualitative methodology and sociology.* Aldershot: Gower.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research: Techniques and procedures for developing grounded theory.* Thousand Oaks, CA: Sage.

Suchman, L. (1995). Making work visible. *Communications of the ACM* [Special issue], *38*(9), 56–64.

von Hippel, E. (1976). The dominant role of users in the scientific instrument innovation process. *Research Policy, 5*(3), 212–239.

Wyche, S. P., Oreglia, E., Ames, M., Hoadley, C. Johri, A., Sengers, P., & Steinfield, C. (2012). Learning from marginalized users: reciprocity in HCI4D. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work Companion* (pp. 27–28). New York: ACM.

8

User Modeling

Pradipta Biswas and Mark Springett

Introduction

The concept of user modeling originated at almost the same time as human-computer interaction. In their seminal book, *The Psychology of Human Computer Interaction*, Card, Moran, and Newell proposed the Model Human Processor, which is probably still the most popular user model in HCI. However, the idea of simulating operators' performance to optimize use of hardware gained importance during the Second World War in the context of designing new military hardware. Computational cognitive psychology then provided the much-needed models of vision, neurons, rapid aiming movements, and so on. With the advent of digital electronics, and more recently the World Wide Web, modeling users has become more complex. In fact, it is now quite difficult to standardize a universal definition of user model that can accommodate all research under the banner of user modeling.

In this chapter, we shall consider any machine-readable representation of its operators as a user model, and we shall summarize conflicting issues about user modeling in the next section. The subsequent sections of this chapter summarize the state of the art in terms of modeling cognition, perception, and motor action, and highlight an application of user modeling to facilitate human machine interaction in a device- and application-agnostic way. Finally, we discuss recent efforts from the European Commission and the International Telecommunication Union (ITU) to standardize user modeling for different applications.

The following sections present a few representative works that were used in HCI either as user models or for simulating users' performance, knowledge or competence. We followed the notion of a model human processor and classified the models into three main categories as cognitive, perception, and movement models.

The Wiley Handbook of Human Computer Interaction, Volume 1, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

Modeling Cognition

Cognition has a number of imprecise and hard-to-measure aspects. As described in the previous section, the aims of cognitive models vary from predictive to exploratory. Predictive models are attractive as a quick way of evaluating user interfaces by estimating task completion times but modern thinking in understanding cognition sees a different set of roles for theory and models of cognition, where the power is not in prediction as such but rather in their explanatory power. Description of the phenomena of interest allows us to acquire a conceptual framework for understanding (Halverson, 2002). Predicting or exploring cognition on its own is often not adequate and we need a different set of models to account for the environment. The following sections describe a set of cognitive models, classifying them based on their prospective application, although it may be noted that a predictive model can also be used as an exploratory model and vice versa.

Predictive models: The model human processor and GOMS model

The model human processor is one of the earliest models applied to HCI. It models human information processing in terms of inputs, outputs, storage, and processing. The model divides into three subsystems, perceptual, cognition, and motor. Two practically applicable versions of these models are the keystroke level model (KLM) and GOMS (goals, operators, methods, and selection rules), which predict error-free expert performance. The GOMS model was inspired by the GPS system (Newell & Simon, 1995) developed by Newell. It assumes that people interact with a computer to achieve a goal by selecting a method, which consists of a sequence of basic operations. The GOMS model enables a designer to simulate the sequence of actions of a user while undertaking a task by decomposing the task into goals and subgoals (John & Kieras, 1996). There are many variations of the original GOMS model.

The keystroke level model (KLM) (Card, Moran & Newell, 1983) simplifies the GOMS model by eliminating the goals, methods, and selection rules, leaving only six primitive operators. They are:

- pressing a key;
- moving the pointing device to a specific location;
- making pointer drag movements;
- performing mental preparation;
- moving hands to appropriate locations, and
- waiting for the computer to execute a command.

The duration of these six operations have been empirically determined. The task completion time is predicted by the number of times each type of operation must occur to accomplish the task.

Kieras developed a structured language representation of GOMS model, called NGOMSL (Natural GOMS Language) (Kieras, 1994). Originally, it was an attempt to represent the content of a cognitive complexity theory (CCT) model (Johnson, 1992) at a higher level of notation. Cognitive complexity theory is a rule-based system developed by Bovair, Kieras, and Polson (1990) to model the knowledge of users of an interactive computer system. In NGOMSL, the methods of the original

GOMS model are represented in terms of production rules of the CCT model. Kieras Wood, Abotel, and Hornof (1995) also developed a modeling tool, GLEAN (GOMS Language Evaluation and Analysis), to execute NGOMSL. It simulates the interaction between a simulated user with a simulated device for undertaking a task.

John and Kieras (1996) proposed a new version of the GOMS model, called CPM-GOMS, to explore the parallelism in users' actions. This model decomposes a task into an activity network (instead of a serial stream) of basic operations (as defined by KLM), and predicts the task completion time based on the critical path method.

Cognitive complexity theory (CCT) uses similar notions of goal structures to GOMS, characterizing them as held in long-term memory and accessed during task performance. This introduces a notion of the way that computer interfaces restructure tasks, and the demands that are placed on users. The restructuring of tasks familiar from the nondigital world is a significant issue in HCI design as users bring their high-level device independent task models to interaction. Complexity in CCT is seen as measurable by the number of production rules required in its performance. It is measured by the number of productions required in the user's notion of how to perform an individual task. This is then compared to the actual number of rules required to perform the task on a particular device.

A number of established predictive models assume the persistence of motivation, that users' commitment to task completion is dependent only on their ability to find a pathway to successful completion limiting the scope of predictive models. The next section presents a set of theories that tried to simulate human cognition and have been used to explain human computer interaction.

Explanatory models

Cognitive architectures Cognitive architectures are a class of models intending to simulate every aspect of human cognition. Allen Newell (1990) developed the SOAR (state operator and result) architecture as a possible candidate for his unified theories of cognition. According to Newell (1990) and Johnson-Laird (1988), the vast variety of human response functions for different stimuli in the environment can be explained by a symbolic system. So the SOAR system models human cognition as a rule-based system and any task is carried out by a search in a problem space. The heart of the SOAR system is its chunking mechanism. Chunking is "a way of converting goal-based problem solving into accessible long-term memory (productions)" (Newell, 1990). It operates in the following way. During a problem-solving task, whenever the system cannot determine a single operator for achieving a task, and thus cannot move to a new state, an impasse is said to occur. An impasse models a situation where a user does not have sufficient knowledge to carry out a task. At this stage SOAR explores all possible operators and selects the one that brings it nearest to the goal. It then learns a rule that can solve a similar situation in future. Laird and colleagues successfully explained the power law of practice through the chunking mechanism.

However, there are certain aspects of human cognition (such as perception, recognition, motor action) that can better be explained by a connectionist approach than a symbolic one. It is believed that initially conscious processes control our responses to any situation while after sufficient practice, automatic processes are in charge for the same set of responses (Hampson & Morris, 1996). Lallement and Alexandre (1997) have classified all cognitive processes into synthetic or analytical processes. Synthetic operations are concerned with low-level, nondecomposable, unconscious, perceptual tasks. In contrast, analytical operations signify high-level, conscious, decomposable, reasoning tasks. From the modeling point of view, synthetic operations can be mapped on to connectionist models while analytic operations correspond to symbolic models. Considering these facts, the ACT-R system (adaptive control of thought—Rational, Anderson & Lebiere, 1998) does not follow the pure symbolic modeling strategy of the SOAR; rather it was developed as a hybrid model, which has both symbolic and subsymbolic levels of processing. At the symbolic level, ACT-R operates as a rule-based system. It divides the long-term memory into declarative and procedural memory. Declarative memory is used to store facts in the form of "chunks" and the procedural memory stores production rules. The system works to achieve a goal by firing appropriate productions from the production memory and retrieving relevant facts from the declarative memory. However, the variability of human behavior is modeled at the subsymbolic level. The long-term memory is implemented as a semantic network. Calculation of the retrieval time of a fact and conflict resolution among rules is done based on the activation values of the nodes and links of the semantic network.

The EPIC (Executive-Process/Interactive Control) (Kieras & Meyer, 1990) architecture incorporates separate perception and motor behavior modules in a cognitive architecture. It mainly concentrates on modeling the capability of simultaneous multiple task performance of users. It also inspired the ACT-R architecture to install separate perception and motor modules and developing the ACT-R/PM system. A few examples of their usage in HCI are the modeling of menu-searching and icon-searching tasks (Byrne, 2001; Hornof & Kieras, 1997).

The CORE system (Constraint-based Optimizing Reasoning Engine) (Eng et al., 2006; Tollinger et al., 2005) takes a different approach to model cognition. Instead of a rule-based system, it models cognition as a set of constraints and an objective function. Constraints are specified in terms of the relationship between events in the environment, tasks, and psychological processes. Unlike the other systems, it does not execute a task hierarchy; rather prediction is obtained by solving a constraint satisfaction problem. The objective function of the problem can be tuned to simulate the flexibility in human behavior.

Additional cognitive architectures exist (such as Interactive Cognitive Subsystems, Apex, DUAL, CLARION, etc.), but they are not yet as extensively used as the previously discussed systems.

Grammar-based models Another approach is the use of grammar models such as task-action grammars, which model memory for the interface language of computer systems. This assumes that task knowledge structures are held in long-term memory and that successful interaction involves yoking of these structural models to the task structure as represented at the interface.

The task-action grammar (Payne & Green, 1986) and task-action language (Reisner, 1981) simulate an interaction in the form of grammatical rules. As for example, task-action language models:

- operations by terminal symbols;
- interaction by a set of rules;
- knowledge by sentences.

This type of modeling is quite useful to compare different interaction techniques. However, they are more relevant to model knowledge and competence of a user than performance.

Environment models The models described above focus on individual cognition and the structures of tasks, or in the head of the individual and in the environment. What these models do not consider is the role of the environment. Norman's Theory of Action (Norman, 1988) describes action as essentially a process of mapping between users prior task knowledge and physical representations on a device. This represents a notion of environmentally led action. The model has seven stages, starting with the formation of a goal. This is followed by the forming of intentions, i.e., specific statements of what has to be done to satisfy the goal. This is in turn followed by the specification of an action sequence. This implies the scanning of the environment for features (or affordances) that can be identified as instrumental to a desired action. The execution of action follows this. The remaining three stages of action are collectively the evaluation phase. Perceiving the state of the world is the initial recognition that a state change has occurred. In the next phase, "interpreting the state of the world," the user tries to make sense of the change, and in the "evaluating the outcome" phase the state change is assessed in terms of progress towards goal satisfaction. So the state change may in some way be unexpected, but if progress is suggested then it is satisfactory. Unexpected/unsatisfactory changes may, if understood, contribute to learning as they may demonstrate an undiscovered system principle.

Norman's model of action provides a tool for thought that supports understanding of the way in which environmentally led action proceeds. It also implies that learning in unfamiliar environments takes place by synthesizing newly encountered environment features with prior knowledge such as structural and procedural task knowledge, and using recognition of familiar names and real-world metaphors.

Rasmussen provides the skills, rules, knowledge (SRK) framework, which is useful in completing the picture both of the nature of display-based interaction and ways in which the user's internal resources synthesize with the environment. It distinguishes skill-based, rule-based, and knowledge-based processing levels. Skill-based processing is automatic nonconscious execution of action most typically associated with experts performing familiar tasks. Rule-based processing requires more working memory capacity using if-then rules to specify action. Knowledge-based processing is reasoning where prior experience is not an available resource. This framework can be usefully applied in understanding end user systems, where the rapid acquisition of a rule base and in turn a skill base to apply to action can be affected through appropriate metaphors, consistent layout and behavior, and system feedback on user action.

Other Notable Approaches The external cognition model (Scaife & Rogers, 1996) takes the principle of understanding cognition as an interplay between internal resources (in the head of the user) and external resources (displays, artefacts, general environmental features). External representations are analyzed in terms of a "cognitive interplay." Three principles underpin this. One is computational offloading, where a good design is one that optimally reduces cognitive effort for its user.

Distributed cognition refers both to interplay of knowledge resources between an individual and external representations but also social distribution, for example in group working environments. An example of this is the resources model, which models information structures and information strategies.

A further class of models consider cognition as embodied. Rather than modeling cognition in isolation from the physical body, these accounts of cognition characterize cognition as inextricably linked with physicality. This represents a fundamental shift in terms of our analysis of behavior, reasoning, and meaning making. This is a relatively underdeveloped area applied to novel technologies such as ubiquitous computing and shared cognitive spaces. An example of theory newly applied in this area is proxemics (Greenberg et al., 2010), where spatial aspects of human behavior and interaction are analyzed.

Modeling Visual Perception

Human vision has been investigated in many ways over the years. The Gestalt psychologists in the early 20th century pioneered an interpretation of the processing mechanisms for sensory information (Hampson & Morris, 1996). Later the Gestalt principle gave birth to the top down or constructivist theories of visual perception. According to this theory, the processing of sensory information is governed by existing knowledge and expectations. On the other hand, bottom-up theorists suggest that perception occurs by automatic and direct processing of stimuli (Itti & Koch, 2001). Considering both approaches, present models of visual perception incorporate both top down and bottom up mechanisms (Neisser, 1976). This is also reflected in recent experimental results in neurophysiology (Reynolds & Desimone, 1999).

Knowledge about theories of perception has helped researchers to develop computational models of visual perception. Marr's model of perception is the pioneer in this field (Marr, 1980) and most of the other models follow its organization. In recent years, a plethora of models have been developed (e.g. ACRONYM, PARVO, CAMERA, and so on (Rosandich, 1997)), which have also been implemented in computer systems. The working principles of these models are based on the general framework proposed in the analysis-by-synthesis model of Neisser (1976) and also quite similar to the feature integration theory of Triesman. It mainly consists of the following three steps:

- **Feature extraction:** As the name suggests, in this step the image is analyzed to extract different features such as color, edge, shape, and curvature. This step mimics neural processing in the V1 region of the brain.
- **Perceptual grouping:** The extracted features are grouped together mainly based on different heuristics or rules (e.g. the proximity and containment rule in the CAMERA system, rules of collinearity, parallelism and terminations in the ACRONYM system). Similar types of perceptual grouping occur in the V2 and V3 regions of the brain.
- **Object recognition:** The grouped features are compared to known objects and the closest match is chosen as the output.

In these three steps, the first step models the bottom-up theory of attention whereas the last two steps are guided by top-down theories. All of these models aim to recognize objects from a background picture, and some of them have been proved successful at recognizing simple objects (such as mechanical instruments). However, they have not demonstrated such good performance at recognizing arbitrary objects (Rosandich, 1997). These early models do not operate at a detailed neurological level. Itti and Koch (2001) present a review of computational models, which try to explain vision at the neurological level. Itti's pure bottom-up model (Itti & Koch, 2001) even worked in some natural environments, but most of these models are used to explain the underlying phenomena of vision (mainly the bottom up theories) rather than prediction. As an example of a predictive model, the VDP model (Daly, 1993) uses image processing algorithms to predict retinal sensitivity for different levels of luminance, contrast and so on. Privitera and Stark (2000) also used different image processing algorithms to identify points of fixations in natural scenes; however, they do not have an explicit model to predict eye movement trajectory.

In the field of human-computer interaction, the EPIC (Kieras & Meyer, 1990) and ACT-R (Anderson & Lebiere, 1998) cognitive architectures have been used to develop perception models for menu-searching and icon-searching tasks. Both the EPIC and ACT-R models (Byrne, 2001; Hornof and Kieras, 1997) are used to explain the results of Nielsen's experiment on searching menu items, and found that users search through a menu list in both systematic and random ways. The ACT-R model has also been used to find out the characteristics of a good icon in the context of an icon-searching task (Fleetwood & Byrne, 2002, 2006). However, the cognitive architectures emphasize modeling human cognition and so the perception and motor modules in these systems are not as well developed as the remainder of the system. The working principles of the perception models in EPIC and ACT-R/PM are simpler than the earlier general purpose computational models of vision. These models do not use any image processing algorithms (Fleetwood & Byrne, 2002, 2006; Hornof & Kieras, 1997). The features of the target objects are manually fed into the system and they are manipulated by handcrafted rules in a rule-based system. As a result, these models do not scale well to general-purpose interaction tasks. It will be hard to model the basic features and perceptual similarities of complex screen objects using propositional clauses. Modeling of visual impairment is particularly difficult using these models. For example, an object seems blurred in a continuous scale for different degrees of visual acuity loss and this continuous scale is hard to model using propositional clauses in ACT-R or EPIC. Shah, Rajyaguru, St. Amant, and Ritter (2003) have proposed the use of image-processing algorithms in a cognitive model but they have not published any result about the predictive power of their model yet. Approaches based on image processing have concentrated on predicting points of fixations in complex scenes while researchers in HCI mainly try to predict the eye-movement trajectories in simple and controlled tasks. There has been less work on using image processing algorithms to predict fixation durations and combining them with a suitable eye-movement strategy in a single model. The EMMA model (Salvucci, 2001) is an attempt in that direction but it does not use any image-processing algorithm to quantify the perceptual similarities among objects.

Biswas and Robinson (2009) proposed a perception model that takes a list of mouse events, a sequence of bitmap images of an interface and locations of different objects in the interface as input, and produces a sequence of eye movements as output. The model is controlled by four free parameters: distance of the user from the screen, foveal angle, parafoveal angle, and periphery angle (Figure 8.1). The default values of these parameters are set according to the EPIC architecture. Biswas and Robinson's model follows the "spotlight" metaphor of visual perception. We perceive something

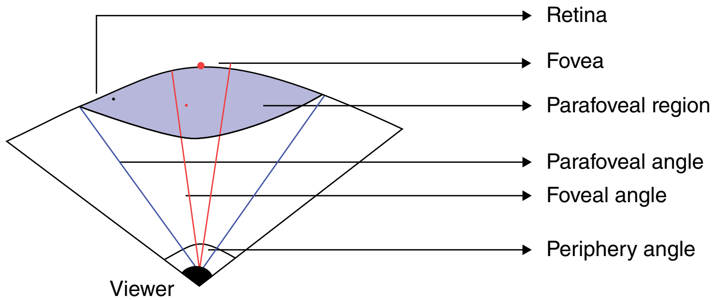


Figure 8.1 Foveal, parafoveal, and peripheral vision.

on a computer screen by focusing attention at a portion of the screen and then searching for the desired object within that area. If the target object is not found we look at other portions of the screen until the object is found or the whole screen is scanned. The model simulates this process in three steps.

1 Scanning the screen and decomposing it into primitive features.
2 Finding the probable points of attention fixation by evaluating the similarity of different regions of the screen to the one containing the target.
3 Deducing a trajectory of eye movement.

Biswas and Robinson's (2009) perception model represents a user's area of attention by defining a focus rectangle within a certain portion of the screen. The area of the focus rectangle is calculated from the distance of the user from the screen and the periphery angle (distance × tan(periphery angle/2), Figure 8.1). If the focus rectangle contains more than one probable target (whose locations are input to the system) then it shrinks in size to investigate each individual item. Similarly, in a sparse area of the screen, the focus rectangle increases in size to reduce the number of attention shifts. The model scans the whole screen by dividing it into several focus rectangles, one of which should contain the actual target. The probable points of attention fixation are calculated by evaluating the similarity of other focus rectangles to the one containing the target. We know which focus rectangle contains the target from the list of mouse events that was input to the system. The similarity is measured by decomposing each focus rectangle into a set of features (color, edge, shape, etc.) and then comparing the values of these features. The focus rectangles are aligned with respect to the objects within them during comparison. Finally, the model shifts attention by combining different eye movement strategies. The model can also simulate the effect of visual impairment on interaction by modifying the input bitmap images according to the nature of the impairment (like blurring for visual acuity loss, changing colors for color blindness). Figure 8.2 shows the actual and predicted eye movement paths (green line for actual, black line for predicted) and points of eye-gaze fixations (overlapping green circles) during a visual search task. The figure shows the prediction for a protanope (a type of color blindness) participant and so the right-hand figure is different from the left-hand one as the effect of protanopia was simulated on the input image. More details on calibration and validation of this model have been described in two different papers (Biswas & Robinson, 2008, 2009).

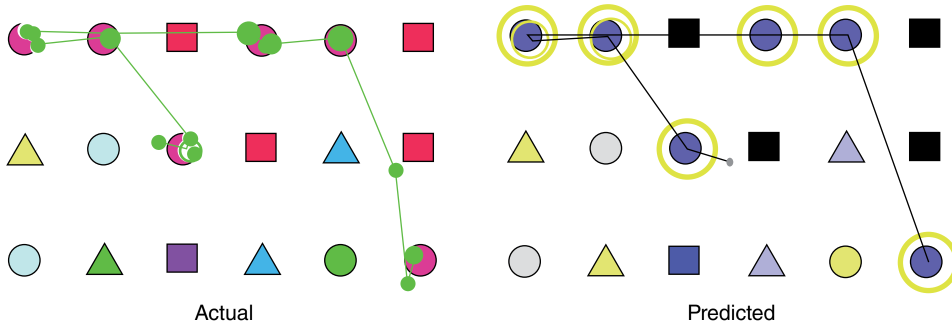


Figure 8.2 Output from a model simulating visual perception of people with visual impairment.

Modeling Movement

Most existing applications in modern electronic devices are based on graphical user interfaces and pointing tasks form a significant part of human machine interaction in those graphical user interfaces. Fitts' law (Fitts, 1954) and its variations are widely used to model pointing as a sequence of rapid aiming movements. Fitts' law (Fitts, 1954) predicts the movement time as a function of the width and distance to the target. This law is found to be very robust and works in many different situations, including in space and under water. Although Fitts proposed his law in 1954, in the late 18th century Woodworth analyzed speed-accuracy tradeoff in rapid aiming movements with or without visual feedback and proposed the existence of an initial impulse and current-control phase in rapid aiming movement. The nature of rapid aiming movement and derivation of Fitts' law (or why it works) was debated among psychologists. Crossman and Goodeve explained Fitts' law in terms of visual feedback and corrective movements while Schimdt explained Fitts' law in terms of neuro-motor impulse. Meyer proposed a generalized equation of predicting movement time for rapid aiming movement, which accommodates Fitts' law as a special case supporting both Crossman and Schimdt's models. The present book dedicates a whole chapter to Fitts' law and its application in HCI.

However the application of Fitts' law for people with motor impairment is debatable. Motor-impaired users only conform to Fitts' law when the task is very simple and thus requires less coordination between vision and motor action (Smits-Engelsman, 2007) or there are other sensory cues besides vision (Gajos, Wobbrock, & Weld, 2007). Existing works in accessible computing mainly point out the problems faced by disabled people in pointing or typing (Keates & Trewin, 2005; Keates, Trewin, & Paradise, 2005; Trewin & Pain, 1999) but there is not much reported work on quantitative analysis of effect of different impairments on pointing performance. Previous works are mainly based on description of diseases or self-report on the level of disability (Gajos, Wobbrock, & Weld, 2007; Hwang, 2002; Trewin & Pain, 1999). A few researchers looked at the reaction time of users—e.g. the Single Switch Performance Test (SSPT), but there is not much reported work on objective evaluation of human factors relating them with human computer interaction (HCI) parameters. Among a few notable attempts, Gajos et al. (2007) found that different combinations of functions involving distance and width of target can predict movement time for different types of mobility impairment. Keates, Clarkson, and Robinson (2000) measured the difference between able-bodied and motor-impaired users with respect to the model human processor (MHP) (Card, Moran, & Newell, 1983) and motor-impaired users were found to have a greater motor action time than their able-bodied counterparts.

In the field of ergonomics, Laursen, Jensen, and Ratkevicius (2001) investigated differences in muscle activities in shoulder, neck, and forearm during pointing, and unsurprisingly concluded that motor impairment demands more motor activity. However they did not try to correlate any pointing parameter with human factors. Smits-Engelsman (2007) found active range of wrist significantly correlate with Fitts' law (Fitts, 1954) constants in pointing tasks for children with congenital spastic hemiplegia. Though the measurement technique is promising, it is yet to be standardized through reliability analysis (test-retest values) like techniques used by ergonomist and occupational therapists. How virtual reality gaming gloves affect natural interaction and works with people having severe finger spasm needs to be investigated.

Biswas and Langdon (2012) analyzed pointing tasks for four different input modalities (mouse, trackball, touch pad, and stylus), and investigated how hand strength affects performance for people with and without mobility impairment. It has been found that the velocity of movement is directly proportional to grip strength, active range of motion (ROM) of wrist and the average number of submovements near the source and target significantly correlate ($p < 0.05$) with logarithm of grip strength and range of motion of wrist for all pointing devices. The correlation between hand-strength metrics and pointing performance was in consistence with results obtained by Smits-Engelsman (2007), and Incel, Ceceli, Durukan, Erdem, and Yorgancioglu (2002). A separate study (Coelho, Duarte, Biswas, & Langdon, 2011) involving a completely different set of participants also confirms that users' preferences match the predicted performance using the model. In particular, users with age-related motor impairment preferred centrally organized targets than peripheral ones, which are also supported by these models, as centrally organized targets need less distance to traverse than peripheral ones and thus require less movement time. The results were used to develop a set of linear regression models that takes grip strength, range of motion of writ, distance to target and size of target as input and predicts pointing time and average number of submovements for different input devices.

Figure 8.3 shows an example of the output from the model. The blue line shows a sample trajectory of the mouse movement of a motor-impaired user. It can be seen that the trajectory contains random movements near the source and the target. The model also predicts approximate movement time and can be run in addition to the visual impairment simulator. In this particular picture, the effect of the early stage of dry macular degeneration was also simulated resulting distortion in the picture. More details on calibration and validation of this model have been described in a different paper (Biswas & Langdon, 2012).

A comparison (Biswas & Langdon, 2012) of the actual pointing times across devices shows that the mouse is fastest for able bodied users while the touchscreen is fastest for disabled users. It can be attributed to the fact that the number of submovements is less for touchscreen or stylus than mouse or trackball. For disabled users, it seems easier to touch the screen directly rather than operating a pointer using a mouse or trackball. However, it is interesting to note that Biswas and Langdon reported that a few users could not use the touch screen at all as they could not unfold their fingers or had significant tremor, while all of them could use the mouse, trackball, or stylus.

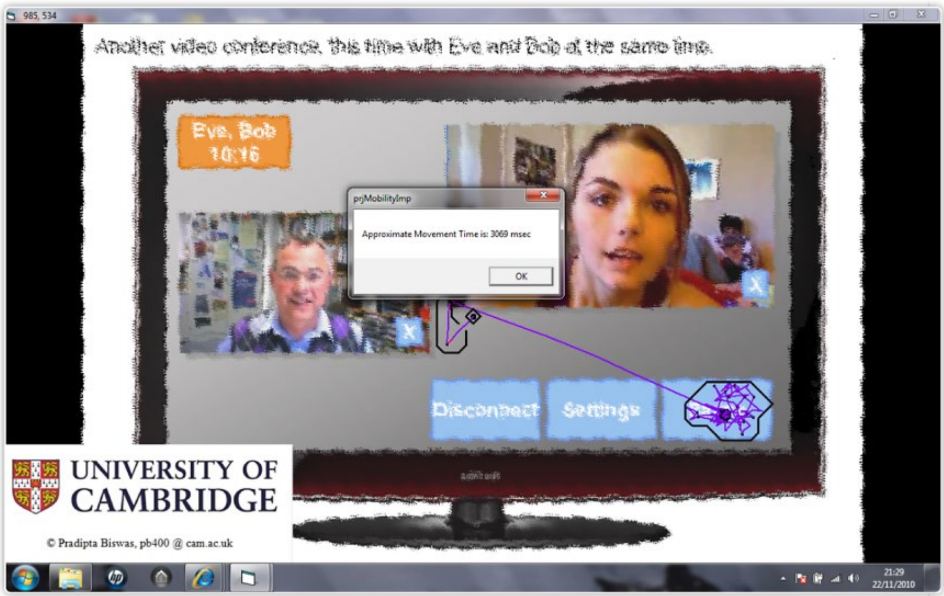


Figure 8.3 Output from a model simulating cursor movement of people with mobility impairment.

Applications of User Modeling

Interface personalization

One of the main applications of user modeling is to adapt user interfaces to facilitate human computer interaction. In the following sections we have presented the inclusive user modeling system—a Web-service-based context-and-device-independent interface personalization system that can adapt both static features of an interface and dynamic interaction based on a stored online user profile.

Stephanidis (1998) and colleagues classified adaptation as static and dynamic adaptation. Static adaptation personalizes an interface before the user starts interacting with it and does not change any feature of the interface during interaction. Dynamic adaptation continuously monitors users' interaction with the system and adjusts features of user interface based on users' interaction. The static adaptation feature of inclusive user modeling systems can automatically adjust font size, color contrast, line, and button spacing of interfaces based on visual acuity, type of color blindness, grip strength, active range of motion of wrist and static tremor of users. The dynamic adaptation part continuously monitor users' mouse movement on screen and based on a least square polynomial curve fitting algorithm, expands users' intended target in the interface. In addition, the inclusive user modeling system

- follows a standardized user profile format specified by a EU cluster and published by the International Telecommunication Union;
- does not propose to change the content of an interface but rather specifies layout parameters, so it is easily integrated with different applications;

- can automatically convert interface parameters (like font size or button spacing) for multiple devices (e.g. TV, computer, laptop, mobile phone and so on) by assuming a viewing distance for different devices and taking the screen resolution as input parameter;
- has investigated details of visual, auditory, and motor functions of humans, and is developed through extensive user trials to relate human factors to interface parameters.

Static adaptation

A user profile creation page (Figure 8.4) takes users' age, sex, height, minimum font size, level of spasm or tremor in hand as input. The Web page also uses an Ishihara color-blindness plate and Amsler grid to detect type of color blindness and distortion

Figure 8.4 User profile creation page.

of vision respectively. The models described in the previous sections on modeling visual perception and movement can predict how a person with visual acuity *v* and contrast sensitivity *s* will perceive an interface or a person with grip strength *g*, and range of motion of wrist (ROMW) *w*, will use a pointing device. This range of values is used in Monte Carlo simulation to develop a set of rules relating users' range of abilities with interface parameters like font size, color contrast, line spacing, default zooming level, and so on. The rule-based system, along with the user, device, and application profiles, are stored in a cloud-based server.

Once users sign up, their profile is stored in a cloud-based server and is accessible to any device or application. The rule-based system selects an appropriate stylesheet (for Web-based systems) or set of interface parameters (for non-Web-based systems) as output. Client applications read data from the user model and sensor network (if they have access to it) and change the font size, font color, line spacing, default zooming level, and so on, by either selecting an appropriate predefined stylesheet or changing parameters for each individual Web page or standalone application. Figure 8.5 shows examples of such integration for Web-based, mobile-phone-based, Android-based and digital-TV-based systems.

Dynamic adaptation

Besides customizing static features of an interface, we have further developed a pointing-facilitation system that reads instantaneous cursor movements and expands

(a)

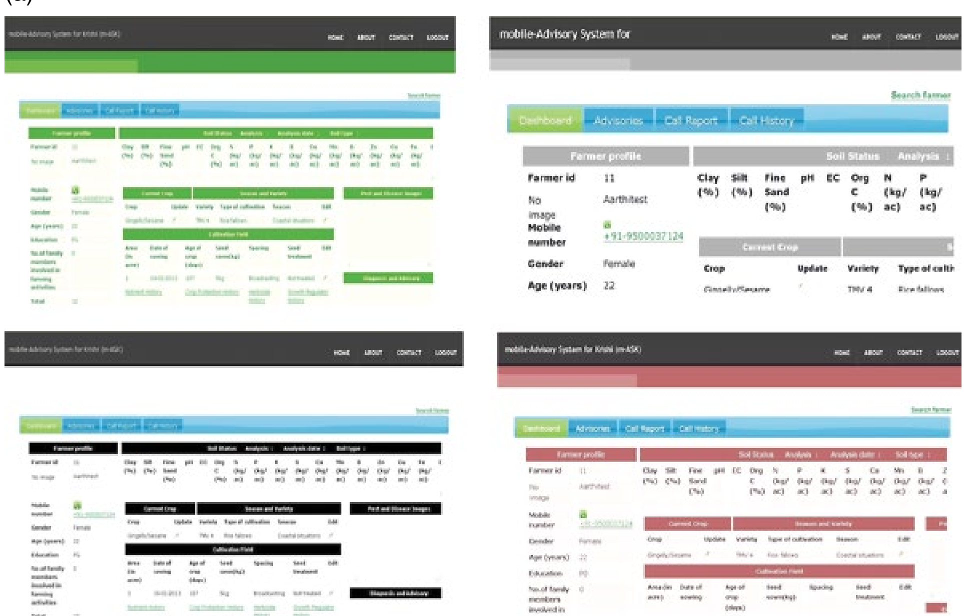


Figure 8.5 Examples of personalized systems. (a) Personalized agri-advisory system (Web based). (b) Personalized agri-advisory system (mobile-phone based). (c) Personalized disaster warning system (Android based). (d) Personalized digital TV framework (set-top box based).

(b)

(c)

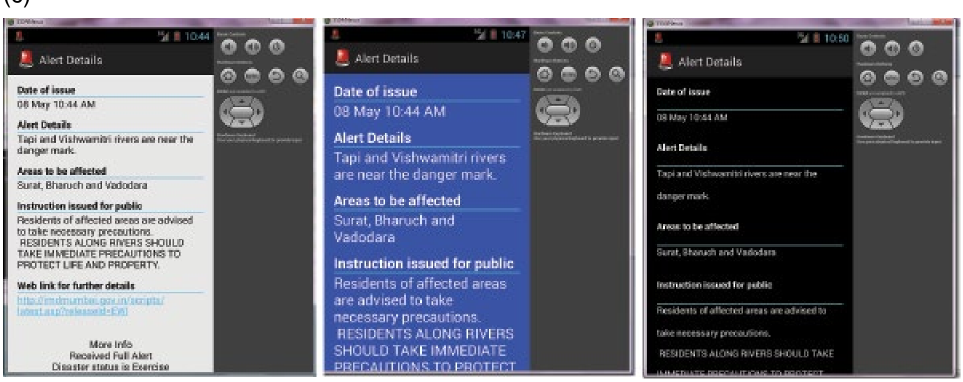


Figure 8.5 (Continued)

(d)

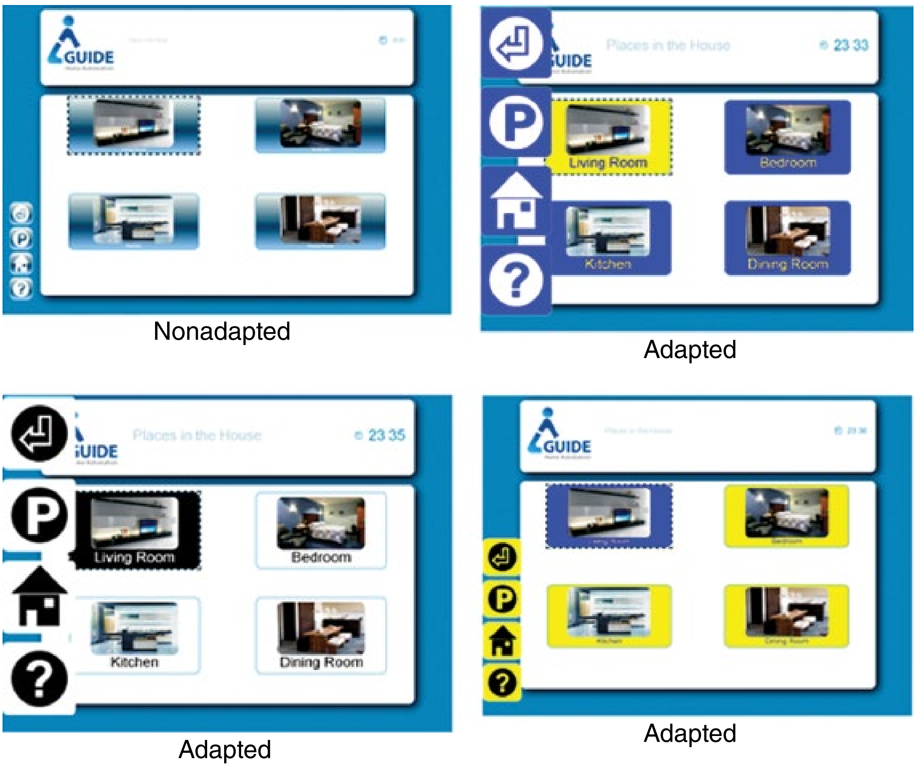


Figure 8.5 (Continued)

onscreen target based on that. The pointing-facilitation system has the following two steps:

1 Smoothing cursor trajectory based on a polynomial algorithm.
2 Scoring a value for each onscreen target and expanding onscreen targets in proportion to their probability of being selected.

The following sections explain these steps in further details.

Smoothing algorithm Previous analysis (Biswas & Langdon, 2012; Keates, Trewin, & Paradise, 2005; Trewin & Pain, 1999) of cursor trajectories of people with hand impairment showed that a cursor movement consists of many submovements. The number of submovements increases near the target when users try to stop the cursor movement. The presence of submovements introduces random jitter in cursor movement. We investigated different smoothing and filtering techniques like Kalman Filter, Bezier Curve fitting, least-square curve fitting for polynomial of degrees from 1 to 8 on cursor trajectories generated by people with different ranges of motor abilities, and based on the quality of fit (R-square and error) we selected a quartic equation to smooth cursor movements. The adaptation system stores a certain number of previous mouse movements to obtain the least squares fit of the last mouse movements creating the smoothed trajectory (Figure 8.6).

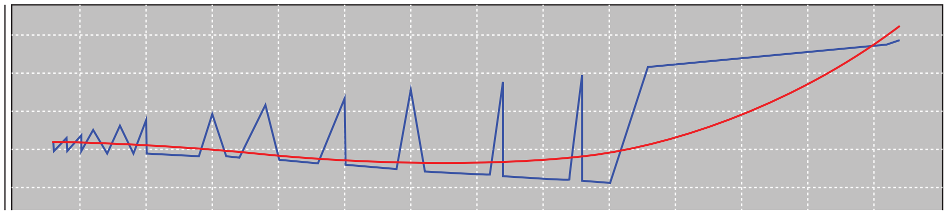


Figure 8.6 An example of polynomial smoothing on a jerky pointing movement. The red line is the smoothed line, and the blue line is the original movement.

Target expansion algorithm After smoothing the cursor trajectory, the adaptation system attempts to predict whether the user is in the ballistic or homing phase of movement (Fitts, 1954) based on instantaneous acceleration of cursor movement, and then uses either of the below methods to calculate the probabilities of selection of targets.

If the cursor movement is in ballistic phase, we assign a score to each target based on bearing of the movement. Upon every mouse movement, the angle between the mouse's direction vector and the target's center are considered, and this angle in radians is added to an array of scores for that target. The direction vector is obtained by interpolating the last three smoothed mouse movements. We only considered movements towards the center of the target. However, it may be prudent to add code to deal with users moving towards an edge of a target in a system where larger buttons are used.

During the homing phase, the score is purely based on the distance to the target in the direction of the cursor, with closer distances having lower scores. When the cursor first reaches a target, its score is reduced to zero, and after that every mouse movement over the target adds a constant value to the score. Finally, we expanded the most probable target (the one that has the highest score) 40% bigger than its original size at a frequency of 7 Hz based on the output of the target scoring algorithm. It may be noted that the frequency of expanding target is reduced from the screen refresh rate to reduce flickering of the targets as the user moves the mouse pointer.

Figure 8.7 shows an example of the adapted interface. The left-hand-side buttons are enlarged (encircled by a gray ellipse) following the pointing facilitation system while the font size and color contrast are adjusted based on the range of abilities of users. The system can be tested at www-edc.eng.cam.ac.uk/~pb400/CambUM/Weather.html

Interaction simulation

User modeling in HCI is primarily used to adapt user interfaces or interaction but it can also be used in design time to evaluate different design alternatives before implementation. One such application of user modeling is in simulating users' interaction. Predictive cognitive models like GOMS or KLM are used extensively to calculate task-completion times of expert users. In the following subsections we presented a couple of case studies of simulating visual perception and cursor movement of people with different ranges of abilities.

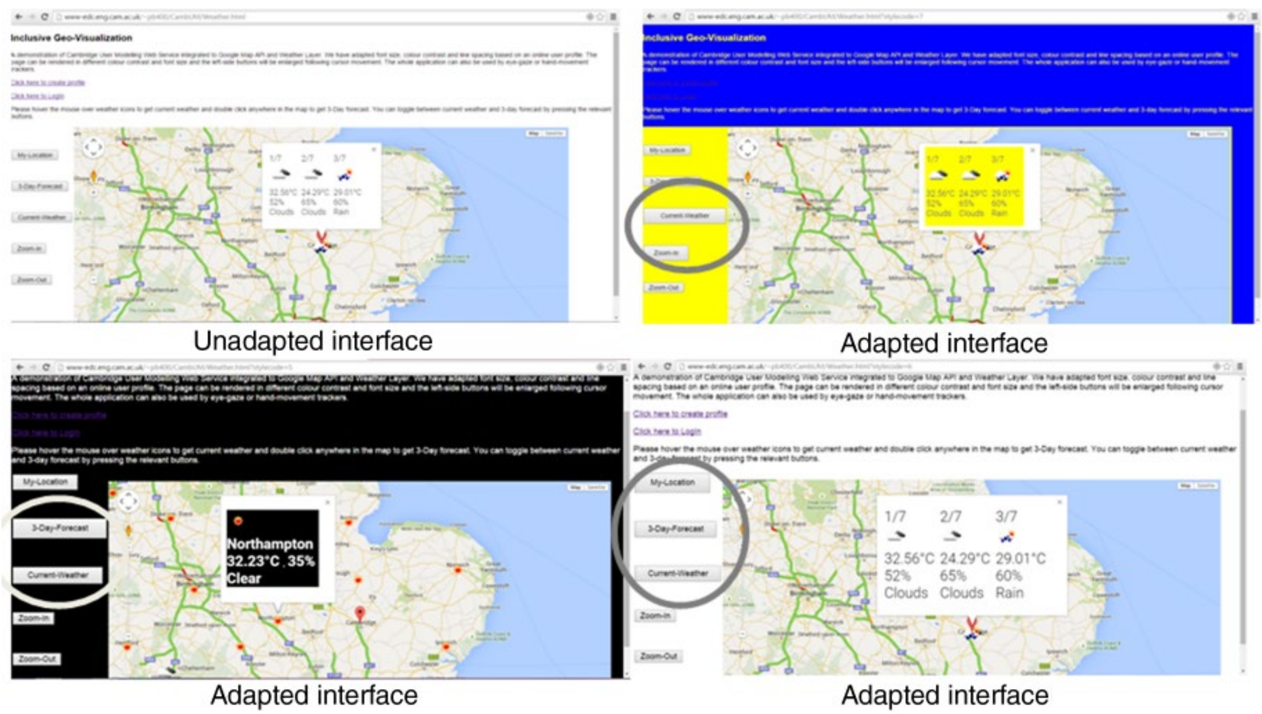


Figure 8.7 Adaptive interfaces demonstrating both static and dynamic adaptation.

Software interface design

The European project GUIDE ("Gentle User Interfaces for Elderly People") created a software framework and design tools, which allowed developers to integrate accessibility and personalization features into their applications efficiently, minimizing intervention with existing development processes and tools. The GUIDE project used a simulator to improve the design of a set of digital TV applications. One such application was the Media Access application intended to be used like an electronic program guide.

The Media Access application was improved after checking simulation results on the first version for user profiles having disabilities, such as age-related shortsightedness and Parkinson's disease. Thus, adjustments were done after analyzing the simulator results found in Figure 8.8 where the white background was seen as too bright, especially for a user with mild visual impairment, or where the focusable arrows were seen as not distinguishable enough. It was therefore decided to make the background darker and the focusable parts were bordered and enlarged when possible.

After doing these modifications, the simulator was reused on the new designs to control the efficiency of the recommendations. These new design simulator results can be seen in Figure 8.9, where a greyish background has been preferred, and the selectable zones have been increased, not only containing the newly bordered arrows but also the full displayed zone.

This last simulation step was perceived as conclusive enough. Therefore no more refinement was decided on the Media Access application.

Nonelectronic product design

The previous example demonstrates the use of the simulator in improving electronic interfaces. However, the simulator can also be used to improve design of physical interfaces. This section explains how the simulator can help product designers to understand effect of different types of visual impairment and provides useful information to improve the design.

The following example demonstrates how designers can decide whether a particular brand of product can be confused with other similar brands by people having mild macular degeneration and red-green color blindness. We start with the following target product shown in Figure 8.10 and compared it with three similar brands. Figure 8.11 shows the simulation of red-green color blindness and early stage of macular degeneration on the target product and three other similar brands. The change in color and blurred images will help designers to visualize the issues with people with visual impairment.

However, we conducted more detailed analysis, and Figure 8.12 and Table 8.1 show the similarity in color and shape features of these products for people with and without visual impairment. We assume a 2-D grid of products as they are arranged in a supermarket or online shopping Web page. We measured the color histogram (an algorithm comparing color features in two images) and shape context (an algorithm compares shapes of two images) coefficients on a scale of 0 to 1 between our target product (marked with a red circle in Figure 8.12) with other similar brands. We have

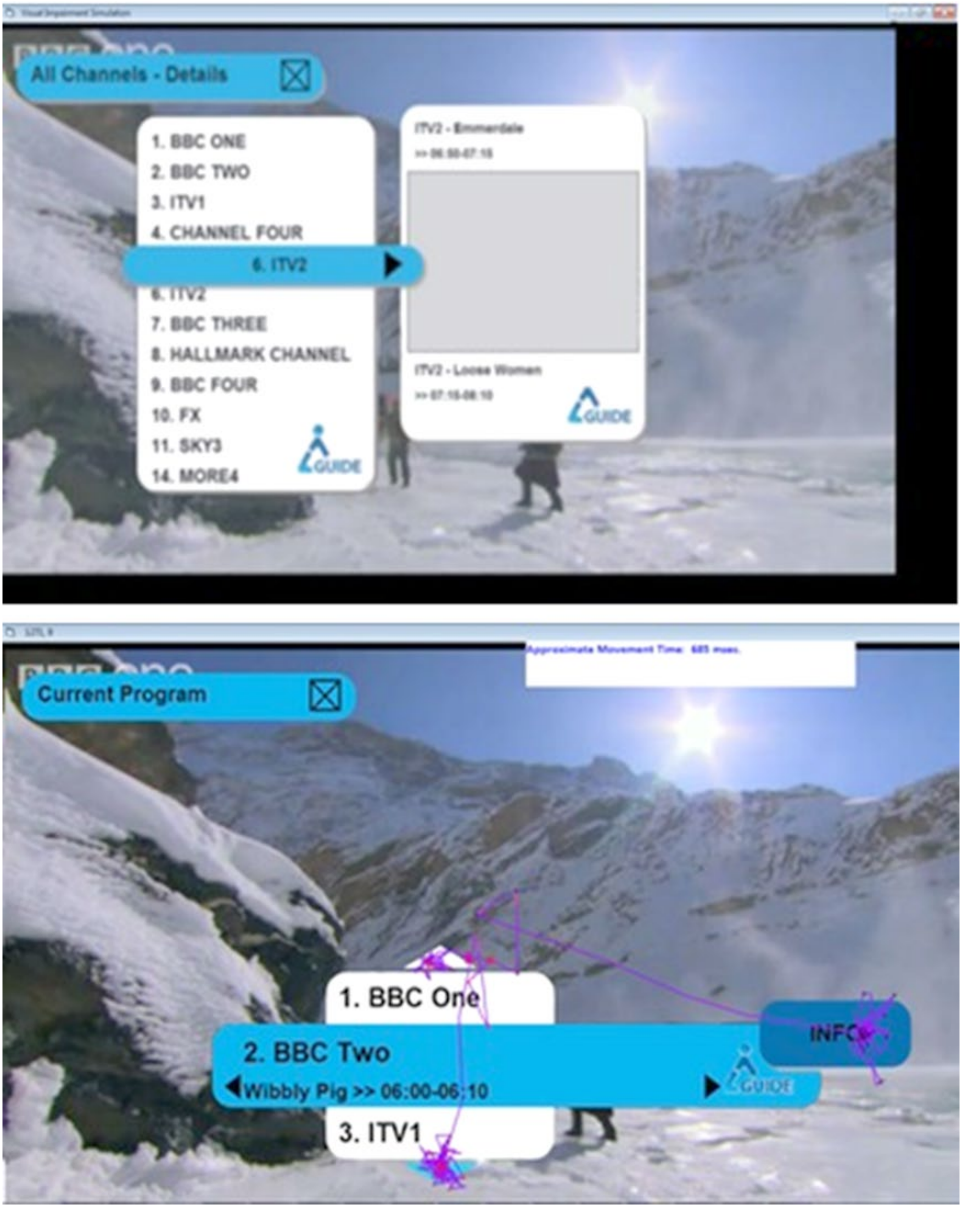


Figure 8.8 Simulator results for mild visual and severe motor impairment on the first version.

put a set of captions below each brand, which all start with the letters "c" and "o." The captions are placeholders for pricing information as it appeared below product pictures and they are purposefully kept visually similar. In Table 8.1, the target brand is colored red and brands that are different in both color and shape are boldfaced. The table also shows there is no significant difference between people with no visual impairment and red-green color blind users for these particular brands. However, for

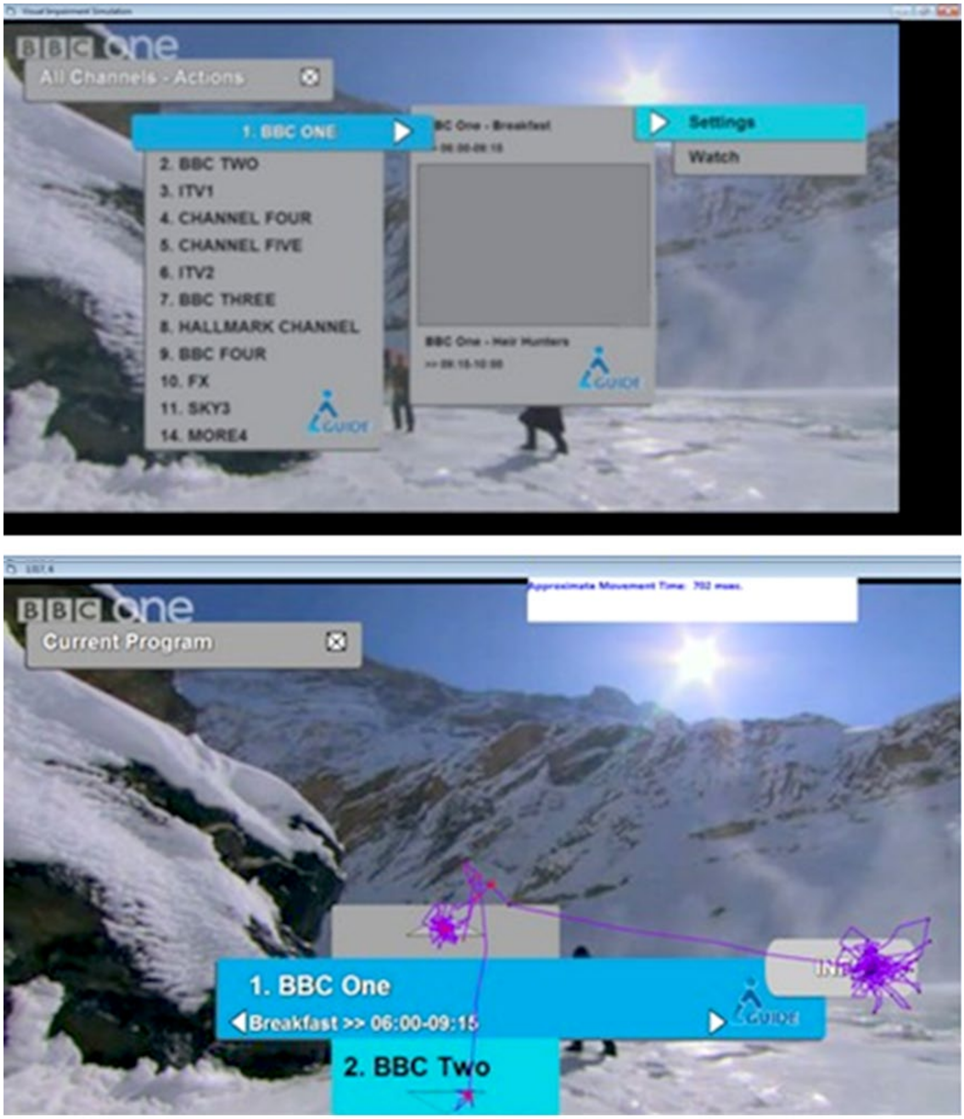


Figure 8.9 Simulator results for mild visual and severe motor impairment on the new design version after the application of the recommendations.

people with macular degeneration, both color histogram and shape context coefficients are reduced, and in particular the color histogram coefficients become similar between the target brand and another similar brand due to blurred and distorted vision. So the simulation suggests that the target brand should have more distinctive color to cover people with blurred and distorted vision. The analysis can be extended to pinpoint color and shape features that can make a target brand look similar to other brands for people with different range of abilities.

Figure 8.10 Target brand.

Figure 8.11 Simulation of color blindness and early stage of macular degeneration.

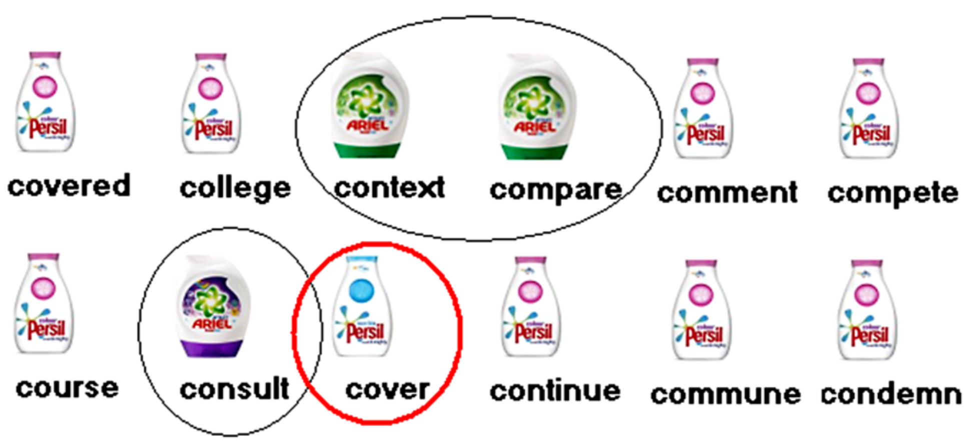


Figure 8.12 A grid of different products.

Table 8.1 Color Histogram and Shape Context Coefficients.

	No visual impairment		*Color blindness*		*Macular degeneration*	
	Color histogram	*Shape context*	*Color histogram*	*Shape context*	*Color histogram*	*Shape context*
1	0.91	0.71	0.89	0.72	0.87	0.40
2	0.91	0.71	0.89	0.72	0.88	0.41
3	**0.80**	**0.57**	**0.81**	**0.57**	**0.79**	**0.26**
4	**0.79**	**0.57**	**0.81**	**0.57**	**0.79**	**0.27**
5	0.91	0.71	0.88	0.72	0.87	0.39
6	0.91	0.71	0.89	0.72	0.87	0.40
7	0.92	0.71	0.90	0.72	0.88	0.40
8	**0.82**	**0.57**	**0.83**	**0.57**	**0.81**	**0.23**
9	**0.96**	**0.78**	**0.95**	**0.78**	**0.89**	**0.65**
10	0.91	0.71	0.89	0.72	0.88	0.40
11	0.91	0.71	0.88	0.72	0.87	0.38
12	0.91	0.71	0.88	0.72	0.88	0.38

Issues with User Modeling

The previous sections of this chapter presented a plethora of models with a particular case study of the "inclusive user model." However, user modeling is still an active research area with the following open questions.

Fidelity

Card, Moran, and Newell's Model Human Processor was a simplified representation of the whole gamut of human-computer interaction, considering perception, cognition, and motor action while Fleetwood and colleagues' model analyzed only eye gaze-tracking movement in the context of viewing icons. Simulation of human activity can range from modeling neuronal response in specific brain region to predicting query terms by analyzing previous search strings. In the context of user modeling, it is often difficult to decide the level of fidelity of the model. Kieras also demonstrated through a case study involving the GOMS model and EPIC architecture that a high-fidelity model does not necessarily signify more accuracy in modeling as each new level of detailed modeling also introduces new error.

Purpose

Why do we need a user model—is it for analyzing a particular type of human computer interaction (for example Fleetwood's work on analyzing visual search in the context of designing icons) or do we need the model to predict human performance like task-completion time or the number of errors in the process of evaluating an interface? The purpose of user modeling is also related to the fidelity issue discussed above. An exploratory model needs to investigate human factors in detail and generally opts for a high-fidelity model like the ACT-R cognitive architecture. Predictive models are more application oriented and can work with engineering approximation (like the timing approximation of KLM) of a high-fidelity model.

Application

Application of a user model also dictates the type and fidelity of the model. Models developed for eLearning or Recommendation systems need not model basic psychology; rather, they can work with more explicit representation of human-computer interaction in the form of previous search queries, semantic models of language, and so on. This type of model is often integrated into an application and is not compatible with other applications except the one it is designed for. However, models like GOMS or CogTool are developed for empirical evaluation of any user interface and investigate human factors in an application-agnostic way. Finally cognitive architectures like ACT-R, EPIC or CORE aim to explain specific types of human behavior even outside the context of HCI, and so using an ACT-R or EPIC model requires a lot more parameter tuning and knowledge about psychology than using a GOMS model.

Representation

User models involving GOMS or cognitive architectures like ACT-R are used to simulate general cognitive performance of users and matches data to individual user through tuning parameters. However, many user models need to predict performance of individual users, for example predicting a query term based on someone's browsing history. These predictive user models needs an efficient way to store user characteristics in the form of a user profile. The content of a user profile varies widely across applications but the increasing number of eCommerce and Web-based adaptive systems created a number of markup languages (ML) to store user profiles. A few examples are UsiXML, EMMA (Extensible Multi-Modal Annotation markup language), MARIA XML and so on. The following section discusses an effort by the European Commission to standardize a generic format for user profile.

Standardization

In 2010, the European Commission took an initiative to standardize user modeling research involving four different EU projects (GUIDE, MyUI, VERITAS, VICON). The consortium was named VUMS (Virtual User Model and Simulation) and mainly aimed to:

- develop a common format of user profile;
- enable sharing of data on user profiles among different projects;
- set up a common ethics format.

The VUMS consortium over the next three years figured out a vocabulary of terms about user model and assembled a super set of users' characteristics relevant for designing accessible car, consumer electronic products to computer software.

Table 8.2 Format of user profile.

Variable name	*Description*	*Data type*
Username	A unique ID of user	String
Password	Log-in credential	String
Age	Age of user in years	Integer
Sex	Sex of user	Integer
Height	Standing height of user	Integer
Volume	Preferred volume of speakers	Double
fontSize	Minimum font size of interface captions	Integer
fontColour	Preferred fore color of buttons	String
cursorSize	Size of cursor	Double
cursorColour	Color of cursor	String
Colour Blindness	Presence and type of color blindness, used to predict color contrast of interface	Integer
Tremor	Presence of tremor or spasm in hand	Integer

The VUMS profile was represented in XML so that it can easily be converted to more specific markup languages like UsiXML or UIML.

However due to the versatility of applications, ranging from automotive environments to digital TV frameworks, the VUMS user profile was too long and complex to be adopted by HCI researchers, who were not much concerned about digital human modeling. In 2014, two focus groups of ITU-T working on audiovisual media accessibility (FGAVA) and smart TV (FG Smart TV) published a subset of VUMS user profiles relevant for HCI. The purposes of the user profiles were:

- personalizing interface layout for software applications;
- adapting electronic content based on users' preferences;
- choosing appropriate access services based on users' needs;
- simulating users' interaction patterns while designing user interfaces.

Table 8.2 furnishes the mandatory part of the user profile.

Conclusion

The changing emphasis from generic models, where a typical user is modeled, to more niche and more personalized models reflects a shift in thinking as to the nature of the user population. In its earlier phases the HCI community tended to consider the user population as a relatively homogeneous group. The only division of this population tended to be into those who computer experts and those who were novices. As HCI has matured as a discipline there is greater recognition of diversity in the population and models have increasingly reflected this. Models for HCI established in the 1980s and 1990s have served as a foundation for developments that reflect both changing perceptions of technology and changing priorities in designing for users.

The notion of the average user is one that was reflected in early user models but, more critically, was also a (partly accidental) characteristic of actual design practice in the field. The utility of models in recent years has partly been to assist designers in understanding and factoring in the requirements of those on the outer edges of the user population. Designers generally find it easy to conceive of a "typical" target user population that is fully able bodied, without impairments or cultural disadvantages. As a result, the user population has often been viewed from the center, with insufficient regard paid to the "nonstandard" user. The change in emphasis, particularly in the modeling of perceptual and motor capacities, is partly due to recognition of changes in the population, and partly due to the expansion in form and purpose of software-based systems.

Demographic changes, most notably the increasingly high proportion of older citizens, mean that the notion of an "average" user is increasingly redundant. Users have a variety of individual needs, and software products should understand and cater for a diverse range of needs, hence the deployment of models in designing for universal access and inclusion. This may be to cater for those who have a registered disability but, increasingly, also for a range of relatively minor impairments that impact upon use of software products. Modeling in support of personalization, customization, and flexibility of use is therefore increasingly important.

A further shift in the focus of user models is that more aspects of the human are seen as relevant to interaction. The initial focus of user models was on aspects of the human cognitive system and physical interaction with devices. This remains a core agenda for HCI. However, this agenda's initial focus was at the desktop and mainly in workplaces. Increasingly software-based systems in the home and in the environment, and systems that support full-body interaction, widen the agenda for user modeling. The next generation of user models in HCI is likely therefore to reflect this wider agenda. Furthermore, some aspects of the human as user that are not available to predictive modeling are a key area of focus for design. Sociality, trust, and other aspects of user experience require approaches that stimulate and influence design thinking rather than specifying rigid parameters.

References

Anderson, J. R., & Lebiere, C. (1998). *The atomic components of thought.* Hillsdale, NJ: Lawrence Erlbaum.

Biswas, P., & Langdon, P. (2012). Developing multimodal adaptation algorithm for mobility impaired users by evaluating their hand strength. *International Journal of Human-Computer Interaction, 28*(9), 576–596.

Biswas, P., & Robinson, P. (2009). Modeling perception using image processing algorithms. In *Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology* (pp. 494–503). Swindon: British Computer Society Learning and Development Ltd.

Bovair, S., Kieras, D. E., & Polson, P. G. (1990). The acquisition and performance of text-editing skill: A cognitive complexity analysis. *Human-Computer Interaction 5*, 1–48.

Byrne, M. D. (2001). ACT-R/PM and menu selection: Applying a cognitive architecture to HCI. *International Journal of Human Computer Studies, 55*(1), 41–84.

Card, S., Moran, T., & Newell, A. (1983). *The psychology of human-computer interaction.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Coelho, J., Duarte, C., Biswas, P., & Langdon P. (2011). Developing accessible TV applications. In *Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 131–138). New York, NY: ACM.

Daly, S. (1993). The visible differences predictor: An algorithm for the assessment of image fidelity. In A. B. Watson (Ed.), *Digital images and human vision* (pp. 179–206). Cambridge, MA: MIT Press.

Eng, K., Lewis, R. L., Tollinger, I., Chu, A., Howes, A., & Vera A. (2006). Generating automated predictions of behavior strategically adapted to specific performance objectives. In *Proceedings of the ACM/SIGCHI Conference on Human Factors in Computing Systems* (pp. 621–630). New York, NY: ACM.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology, 47*, 381–391.

Fleetwood, M. F., & Byrne, M. D. (2002). Modeling icon search in ACT-R/PM. *Cognitive Systems Research, 3*(1), 25–33.

Fleetwood, M. F., & Byrne, M. D. (2006). Modeling the visual search of displays: A revised ACT-R model of icon search based on eye-tracking data. *Human-Computer Interaction, 21*(2), 153–197.

Gajos, K. Z., Wobbrock, J. O., & Weld, D. S. (2007). Automatically generating user interfaces adapted to users' motor and vision capabilities. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology* (pp. 231–240). New York, NY: ACM.

Greenberg, S., Marquardt, N., Ballendat, T., Diaz-Marino, R., & Wang, M. (2010). Proxemic interactions: the new ubicomp? *Interactions, 18*, 42–50.

Halverson, C. A. (2002). Activity theory and distributed cognition: Or what does CSCW need to DO with theories? *Computer Supported Cooperative Work, 11*(1–2), 243–267.

Hampson, P., & Morris, P. (1996). *Understanding cognition*. Oxford: Blackwell Publishers.

Hornof, A. J., & Kieras, D. E. (1997). Cognitive modeling reveals menu search is both random and systematic. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (pp. 107–115). New York, NY: ACM.

Incel, N. A., Ceceli, E., Durukan, P. B., Erdem, H. R., & Yorgancioglu, Z. R. (2002). Grip strength: Effect of hand dominance. *Singapore Medical Journal, 43*(5), 234–237.

Itti, L., & Koch, C. (2001). Computational modeling of visual attention. *Nature Reviews, Neuroscience, 2*, 1–10.

John B. E., & Kieras, D. (1996). The GOMS family of user interface analysis techniques: Comparison and contrast. *ACM Transactions on Computer Human Interaction, 3*, 320–351.

Johnson, P. (1992). *Human computer interaction: Psychology, task analysis and software engineering*. London: McGraw-Hill.

Johnson-Laird, P. A. (1988). *The computer and the mind*. Cambridge, MA: Harvard University Press.

Keates, S., & Trewin, S. (2005). Effect of age and Parkinson's disease on cursor positioning using a mouse. In *Proceedings of the Seventh international ACM SIGACCESS Conference on Computers and Accessibility* (pp. 68–75). New York, NY: ACM.

Keates, S., Clarkson, J., & Robinson, P. (2000). Investigating the applicability of user models for motion impaired users. In *Proceedings of the Fourth International ACM Conference on Assistive Technologies* (pp. 129–136). New York, NY: ACM.

Keates, S., Trewin, S., & Paradise, J. (2005). Using pointing devices: quantifying differences across user groups. In *Proceedings of UAHCI 2005: Third International Conference on Universal Access in Human-Computer Interaction*. Mahwah, NJ: Lawrence Erlbaum Associates.

Kieras, D. (1994). *GOMS Modeling of user interfaces using NGOMSL, conference companion on human factors in computing systems* (pp. 371–372). Boston, MA: ACM.

Kieras, D., & Meyer, D. E.. (1990). An overview of the EPIC architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction, 14*, 391–438.

Kieras, D. E., Wood, S. D., Abotel, K. & Hornof A. (1995). GLEAN: A computer-based tool for rapid GOMS model usability evaluation of user interface designs. In *Proceedings of ACM Symposium on User Interface and Software Technology* (pp. 91–100). New York, NY: ACM.

Lallement, Y., & Alexandre, F. (1997). Cognitive aspects of neurosymbolic integration. In R. Sun & F. Alexandre (Eds.), *Connectionist-symbolic integration*. London: Lawrence Erlbaum Associates.

Laursen, B., Jensen, B. R., & Ratkevicius, A. (2001). Performance and muscle activity during computer mouse tasks in young and elderly adults. *European Journal on Applied Physiology, 84*, 329–336.

Marr, D. C. (1980). Visual information processing: The structure and creation of visual representations. *Philosophical Transactions of the Royal Society of London, 290*(1038), 199–218.

Neisser, U. (1976). *Cognition and reality*. San Francisco, CA: Freeman.

Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press.

Newell, A., & Simon, H. A. (1995). *GPS, a program that simulates human thought*. Cambridge, MA: MIT Press.

Nilsen E. L. (1992). *Perceptual-motor control in human-computer interaction* (Technical Report No. 37). Ann Arbor, MI: Cognitive Science and Machine Intelligence Laboratory, University of Michigan.

Norman, D.A. (1988). *The design of everyday things*. Cambridge, MA: MIT Press.

Payne, S. J., & Green, T. R. G. (1986) Task-action grammars: A model of mental representation of task languages. *Human-Computer Interaction, 2*, 93–133.

Privitera, C. M., & Stark, L. W. (2000) Algorithms for defining visual regions-of-interests: Comparison with eye fixations. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 22*(9), 970–982.

Reynolds, J. H., & Desimone, R. (1999). The role of neural mechanisms of attention in solving the binding problem. *Neuron, 24*, 19–29, 111–145.

Rosandich, R. G. (1997). *Intelligent visual inspection using artificial neural networks* (1st ed.). London: Chapman & Hall.

Salvucci, D. D. (2001). An integrated model of eye movements and visual encoding. *Cognitive Systems Research, 1*, 201–220.

Scaife, M., & Rogers, Y. (1996). External cognition: How do graphical representations work? *International Journal of Human-Computer Studies, 45*, 185–213.

Shah, K., Rajyaguru, S., St. Amant, R., & Ritter, F. E. (2003). Connecting a cognitive model to dynamic gaming environments: architectural and image processing issues. In *Proceedings of the Fifth International Conference on Cognitive Modeling* (pp. 189–194). Bamberg, Germany: Universitats-Verlag Bamberg.

Smits-Engelsman, B. C. M. (2007). Children with congential spastic hemiplegia obey Fitts' law in a visually guided tapping task. *Journal of Experimental Brain Research, 177*, 431–439.

Stephanidis, C., Paramythis, A., Sfyrakis, M., Stergiou, A., Maou, N., Leventis, A.… Karagiannidis, C. (1998) Adaptable and adaptive user interfaces for disabled users in the AVANTI Project. In S. Triglia, A. Mullery, M. Campolargo, H. Vanderstraeten, & M. Mampaey (Eds.), *Intelligence in services and networks: Technology for ubiquitous telecom services*. Berlin: Springer.

Tollinger, I., Lewis, R. L., McCurdy, M., Tollinger, P., Vera, A., …. Pelton, L. (2005). Supporting efficient development of cognitive models at multiple skill levels: Exploring recent advances in constraint-based modeling. In *Proceedings of the ACM/SIGCHI Conference on Human Factors in Computing Systems* (pp. 411–420). New York, NY: ACM.

Trewin, S., & Pain, H. (1999). Keyboard and mouse errors due to motor disabilities. *International Journal of Human-Computer Studies 50*(2), 109–144.

9

Kids and Design

Mona Leigh Guha and Jerry Alan Fails

Overview and Introduction

In this chapter we focus on the processes used to design technology for children. Other chapters in this handbook address children and HCI in general (Chapters 14 and 16), as well as educational technology (Chapter 38). While those chapters focus more heavily on technology and interface outcomes, in this chapter we focus on the design process, including the evaluation, of children's technologies.

In recent decades, technology use by children has grown dramatically and it appears as though it will only continue to increase (Rideout, 2013). From Web sites, to apps for mobile devices, to robots and beyond, the possibilities for technology for children are virtually limitless. While the technology for children is in itself interesting, valuable, and the end goal of any technology design process, a large portion of our research work focuses on the *design processes* for creating technology for children. The method by which a technology for children is designed not only inevitably influences the end product but it also speaks to the philosophy and values of the designers, as well as potentially impacts the designers themselves. In this chapter, we outline the history of children's roles in design in relation to the design of children's technology and discuss different methods of designing technology for children. We present information regarding some focused research on designing technology for children in special circumstances. We then consider the related field of evaluation of children's technology. We conclude with a look at future trajectories in technology design with children.

Context and History of Designing Technology for Children

For as long as there has been technology for children, there have been processes by which it is designed. Over time, many processes for designing technology for children have evolved (Fails, Druin, & Guha, 2013). Druin (2002) posited a framework for thinking about the different methods of design for children (Figure 9.1), which also parallels history as one moves from the inside of the circle to the outer rim.

The Wiley Handbook of Human Computer Interaction, Volume 1, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

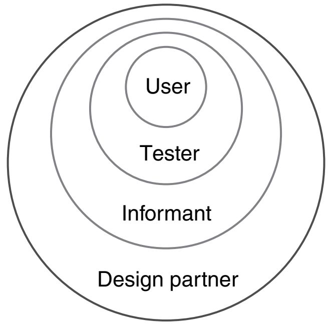


Figure 9.1 Levels of children's involvement in the design process.

This framework provides a context for designing technology for children. It views technology design for children through the lens of the roles that children may play in the design process. We employ this framework to guide the discussion in this chapter.

Context through roles

Beginning from the center of the figure, the first manner in which technology was designed for children considered children as **users** of technology. In this role, children only interact with the technology after it has been fully developed—thus they interact merely as end users of the technology. The interaction between adults and children is typically that adults design technology for children, and children use it after it is fully developed. Adults may then observe children interacting with technology in order to inform future design but typically do not directly ask for the children's opinions. When children are involved as users, adults do the design work. Moving concentrically outward on the diagram, the next role is children as **testers**. When children act as testers, they test at least one iteration of a of a technology before it is deployed to guide adult designers during the design process in order to shape the final technology. Thus, while adults still bear the most responsibility for design, children in the role of tester are able to have impact an impact earlier in the design process than children in the user role.

As we progress to involving children more in the design process represented as the outer circles in the diagram, there is a difference in the manner in which children interact with adults in the design process. At the next levels, children are more integrally involved throughout the technology design process. At the level of **informants**, children begin to give feedback to adult designers throughout the design process. Child informants participate in various design activities at key stages in the design process when the adult designers feel that feedback from children could be informative. Thus, child informants take a more active role than users or testers in that they are involved and their opinions and ideas about the technology are heard in more active ways during the design process. Finally, methods of working with children as **design partners** include children on the technology design team from the beginning to the end of the design process, with children and adults acting as equal members of the team. As implied by the concentric circles in the diagram, children as design partners may occasionally assume the role of users, testers, or informants; however, the end goal is for the children to be integrally involved in the design process. When children are involved in a technology design process as design partners, they can perform the other roles in an iterative and design-focused manner, not as subjects for usability or other testing. As with adults, children should not evaluate technology that they had a voice in creating. Adult and child members of a design partnering team tend to be consistent over years, and the team generally works on several projects throughout the duration of time that they work together. The difference between informant and design partner is that in design partnering, children are involved in a more constant and consistent manner throughout the design process.

History

In HCI, the role of user for children is historically the first role that children played in regard to the design of their technology. This level of involvement began in the 1970s, and was first described in an HCI paper by Malone (Malone, 1982), who observed children as users to determine what made computer games enjoyable. Roles both of testers and informants came about in the 1990s, with informant design coming more to prominence in the late 1990s (Scaife & Rogers, 1999; Scaife, Rogers, Aldrich, & Davies, 1997). Design partnering began in the late 1990s (Druin, 1999) and continued in earnest in the 2000s.

While children are still involved in the design process as users, over the years, children have become more prominently involved in designing their technologies. A recent survey of academic literature reports that children are involved as testers more than users, and that children are frequently involved as informants or design partners (Yarosh, Radu, Hunter, & Rosenbaum, 2011). While many designers still work with children as users and testers in the design process, testing in and of itself is not a method of design. While not explicitly identified as informant or design partnering, iterative testing with users is a form of informant design. In this chapter we focus on the methods of design that involve children in the role of informant or design partners.

Design processes that are more participatory—meaning those that include the end users in a deep manner throughout the design process—began in Scandinavia decades ago. These participatory processes stemmed from demands of the unions that workers' voices be heard in the creation of their work environments, including the technologies that were a part of these environments (Bjerknes, Ehn, & Kyung, 1987; Bødker, Ehn, Sjögren, & Sunblad, 2000). Typically, participatory design methods move toward compromise (Large et al., 2007) and cooperative design strategies (Smith, 2011) on design teams in order to create new technologies. Participatory design (PD) processes are part of the philosophy of user-centered design (UCD) where the user's needs are considered as a central focus, rather than the previously common system-centered approach where the focus was clearly on the technology or system, not on who was using the system (Schuler & Namioka, 1993). Participatory design has been identified as a large space within the HCI community (Muller & Druin, 2012).

There are many reasons that participatory design processes are employed by designers of children's technology. First, participatory design processes can empower the user (Fails et al., 2013; Guha, 2010). Allowing children to have a voice in the design of their technology gives voice to an often underheard group. Technology that is created with users involved in the process is also often more effective for that group, and when it comes to children, can be more innovative. In the next section, we describe some methods of design that have recently been or are currently being used in designing technology for children. In the space of this chapter, we cannot possibly cover all of the methods of design, and variations of methods of design, in use today for those working in technology design for children. Thus, most of the design methods that we present here are more participatory, and would be considered informant or design partnering according to Druin's framework. We provide an overview of different types of participatory methods used to give children a voice in the design of their technologies.

Specific Methods of Designing Technology for Children

A note about "method" and "technique"

In this chapter, we provide an overview of methods used to design technology for children. Before we begin, we need to distinguish our definition of a design method from a design technique. Our definition of "technique" is narrow and refers to the specific activities that are used within a design process. Techniques include activities such as low-tech prototyping and strategies for iterative feedback like sticky note critiquing (Walsh, Foss, Yip, & Druin, 2013). Techniques are practical, are typically short in duration, and often include activities that take a single design session (or less) to complete. Multiple techniques may be used during the course of a design process. A "method" is broader in scope. We use Walsh et al.'s definition of a method as a "collection of techniques used in conjunction with a larger design philosophy" (Walsh et al., 2013, p. 2893). A method includes not only design techniques, but also the overall philosophy of the team. In this chapter, we are presenting information on design methods. For information on specifics techniques used in design processes for children's technology, see Fails, Guha, and Druin (2013).

Many factors may affect a team's decision on which method to choose in working to design technology for children. Certainly personal, professional, and institutional philosophy will affect the decision. There also may be constraints, such as financial and time constraints, in place. Here we provide an overview of some of the methods available to designers.

Cooperative inquiry

Cooperative inquiry is a method of design that evolved from the Scandinavian participatory design (PD) movement. Cooperative inquiry requires children and adults working together on an intergenerational design team to design children's technology. The underlying philosophy of cooperative inquiry is that, while adults may have expertise in specific areas such as computer science, child development, or design, children have expertise in being children today—something that is clearly an expertise unique to them. Therefore, a cooperative inquiry team is not only interdisciplinary in that it includes members from various disciplines; it is also intergenerational in that it includes both adult and child members. Children on a cooperative inquiry design team typically have an age range of years so that technology for many different ages can be designed by the team. For example there have been teams with children ages 7–11, as well as teams for younger children ages 4–6, and older children ages 11–14. All members of the team must bring their unique perspective

Figure 9.2 *Left:* codesign team working together. *Right:* artifacts of design sessions: top, low-tech prototyping, middle, big ideas on white board, bottom, sticky note evaluation.

and expertise to work together in order to produce the best possible technology for children. Typically, cooperative inquiry design teams meet at least weekly throughout the year, and membership both of adults and children stays consistent over the course of a year or many years. Some of the techniques used in cooperative inquiry include low-tech prototyping (bags of stuff), sticky note evaluation (Walsh et al., 2013), and layered elaboration (Walsh et al., 2010). There are extensive publications regarding cooperative inquiry (Druin, 1999; Guha, Druin, & Fails, 2013), and the driving philosophy of cooperative inquiry is to give children a voice by making them partners in the design process. Figure 9.2, illustrates some of the process, and artifacts created through this process.

Over the years, numerous technologies have been developed using the cooperative inquiry design method, including the International Children's Digital Library (http://en.childrenslibrary.org/) (Druin, Bederson, Rose, & Weeks, 2009), an online multilingual library of numerous children's books, as well as BodyVis (Norooz et al., 2015), a wearable shirt created to help children learn about their bodies through technological interaction. Recently, Smith, Iversen, and Lynggaard (2013) have extended the ideas of cooperative inquiry into ecological inquiry, which further emphasizes the ecology of the physical and activity context, as well as how technologies are integrated into a hybrid environment, during the design process. Ecological inquiry accomplishes this through reflection on social practice, physical space, and technology during the design process.

Informant design

As mentioned earlier, informant design is a method of design in which children are consulted at specific points during the design process for their feedback and input. While the techniques used may be substantially the same as those in other methods, the difference is in the timing and consistency with which adult designers consult child designers. In informant design, adult designers determine, on a project-by-project basis, when and how frequently children are included in the design process. Children are brought into the design process when the adult designers feel that their input is necessary or would be particularly helpful with a particular design issue. Many researchers including Scaife, Rogers, Aldrich, and Davies (1997) and Scaife and Rogers (1999) find working with children as informants to be a well-suited role as through this role adult designers can elicit effective feedback from children without requiring a full partnership with the children, which they believe is not possible. Mazzone, Read, and Beale (2008b) extended initial work with informant design with children to accommodate working with teenagers. In this work, teenagers with behavior challenges were included as principal informants in designing a computer program to help teenagers deal with emotions. The philosophy of informant design is to give children a voice in the process by including them at key stages in the design process. This method is broadly used. A couple of examples include designing educational activities in museums (Mazzone et al., 2008b) and designing tangible user interfaces (Xu, Mazzone, & MacFarlane, 2005).

Bonded design

Bonded design (Large, Nesset, Beheshti, & Bowler, 2006; Large, Bowler, Beheshti, & Nesset, 2007) is an intergenerational design method in which children and adults participate together on design teams, but generally for a short and intense amount of time, such as a matter of weeks. The number of projects that a bonded design team works on is therefore generally limited in time and typically focused in project scope. Design techniques in bonded design can be largely the same as those in cooperative inquiry or informant design. The "bond" in bonded design refers to the bond between the children and the adults on the team. Bonded design activities may take place in a school setting, and often the children on the team are of the same age. The philosophy of bonded design is to form a partnership between children and adults, but it is bounded to the duration and topic of a specific project. This method has been used to explore the creation of "Web portals" (now more commonly referred to as search engines) (Large et al., 2006, 2007) and more recently information search in virtual reality libraries (Beheshti & Large, 2011) and an educational tool for students involved in inquiry-based projects (AlGhamdi, Beheshti, Cole, Abuhumed, & Lamoureux, 2013).

Focus groups

In a focus group, a group of users—in this case children—are brought together to give feedback on a specific issue—which in this case is a technology. There are several important factors to consider when conducting focus groups, including group composition, geographic location, and scheduling, creating the right environment, and who will moderate (Gibson, 2007). Typically there are adult facilitators for focus groups, and focus groups can occur at anytime during the design process—from requirements gathering to iterative evaluation to feedback on an end product. While focus groups do not by themselves constitute a complete method of design, they do include children in a participatory way in the design process and are commonly used, especially in industry. While focus groups may seem to be more of a technique than a method, we include focus groups in the discussion of methods due to their versatility and common usage in industry settings. Fisher Price, for example, has used focus groups as a primary source of information to impact developing developing their products (Gutman, 2012). Children's opinions are generally being solicited, and depending on how the focus group is conducted, it can be more tester-oriented, or more informant-oriented (Naranjo-Bock, 2011). Focus groups can be done in a resource-light manner at various times during the design process and can be valuable for gaining children's insights at various times during the design process. Focus group sessions can be conducted in a more tester-oriented manner where users test technology, or more informant-oriented so children help design or suggest improvements to a technology. When multiple focus groups are conducted, different children are generally involved in each session. This adds recruiting overheads, but affords the ability for more children's voices to be heard, although as they interact generally only during one session the bond of trust is not particularly strong between the children and adult designers or researchers. The philosophy of focus groups is to give children a voice during employing a single, short investigative session.

Learner-centered design

As a method, and in philosophy, learner-centered design does not explicitly directly include children in the design process of technology; however, as indicated by the name, in learner-centered design, the focus is on the child as a learner (Soloway, Guzdial, & Hay, 1994; Soloway et al., 1996). Technology designed using learner-centered design includes handheld technologies created to support students *in situ* during research activities by supporting creation and sharing of models and maps (Luchini, Quintana, & Soloway, 2004). Practitioners of learner-centered design consider the act of learning as central to the technology itself and the design process. Learner-centered design originated as a concept to challenge designers to think of end users not only as *users* of technology, but as *learners* as well; thus, the learner in this context can be a child or adult. Learner-centered design is an extension of user-centered design (UCD) where the user is the focus of the design process, not just the system.

Child personas

Child personas refer to the practice of creating narratives of fictional child users for the design team to keep in mind and continually reference as they design technology (Antle, 2006; Antle, 2008). Child personas can be used when there are practical or other barriers to working with real children during the design process, or even in conjunction with children in participatory design. When creating child personas, designers should keep in mind not only to include as much detail as possible, but also to include developmental information about the child personas. This method was used to develop story creating activities for CBC4Kids.ca (Antle, 2003). Some have developed games and health-promoting services by respectively creating child personas through probes (Moser, Fuchsberger, & Tscheligi, 2011), and coconstruction with health-care professionals and parents (Wärnestål, Svedberg, & Nygren, 2014).

This practice is sometimes used in other kinds of settings such as prototypical buyers at a retail store, or relatable prototypes of a person in games, media, or literature. The philosophy and focus is to help the designer relate to the end user of the technology and make sure the persona's needs, in this case children's needs, are fulfilled by the technology that is being designed.

Special Circumstances when Designing with Children

While much technology is created for all children, there are distinct groups of children that have special circumstances. These groups have different needs and require different methods—or at a minimum modifications to existing methods—to allow their voices to be heard clearly throughout the design process. Addressing the needs of some of these more specific groups of children in the design of new technology has already begun, including working with children with special needs, differing ages (from young children to teenagers), and family situations (single-parent homes, etc.). In the following section, we discuss the unique needs of these groups and how they have been addressed in technology design processes.

Special needs

Much work has been done in the area of designing technology for children with special needs. From iPad apps to help children with learning disabilities (Hourcade, Bullock-Rest, & Hansen, 2012), to technology for full-body interaction to encourage children with autism to engage in classrooms (Battacharya, Gelsomini, Perez-Fuster, Abowd, & Rogza, 2015) as well as much work in the area of interactive technology for children with autism (Cramer, Hirano, Tentori, Yeganayan, & Hayes, 2011; Kientz, 2012; Kientz, Goodwin, Hayes, & Abowd, 2013), copious work has been done to create technology to support children with special needs and the adults who work with them in a variety of environments. Technology can provide unique and surprising support for children with special needs.

As with technology for typically developing children, technology for children with disabilities can be designed in a number of ways. Borjessen, Barendregt, Eriksson, and Torgersson (2015) found, through a review of literature, that high-functioning children with autism are the most likely of special needs children to be included in design processes, and that design processes with children with special needs often include a more in-depth adult role, more structure, and less direct input from the children. Frauenberger and his colleagues have been instrumental in considering how children with special needs, particularly children with autism, can be involved in the technology design process, arguing that input from children with disabilities is invaluable as nondisabled designers do not know what it is to experience life with a disability (Frauenberger, Good, & Alcorn, 2012). Methods and techniques including the IDEAS framework (Benton, Johnson, Brosnan, Ashwin, & Grawemeyer, 2011; Benton, Johnson, Ashwin, Brosnan, & Grawemeyer, 2012) and the ECHOES technique (Frauenberger, Good, Alcorn, & Pain, 2012) have been created to support including children with disabilities in the design of their technology. Foss et al. (2013)

extended work in cooperative inquiry to working with boys ages 11–12 with learning differences, and found that cooperative inquiry could be used with this population with modifications such as maintaining a high adult to child ratio and giving multimodal instructions. Mazzone, Read, and Beale (2008a) also worked with slightly older children, in this case 12–15 years old, who had behavioral difficulties and asked these teenage informants to help design computer software to help manage emotions, and offered suggestions for working with children with special needs including keeping groups small and tasks well-defined and small.

Differing ages

When technology is designed for children, the designers must keep in mind the developmental levels of the children. A technology designed for a preschooler can and should look and function in a very different manner from one designed for a teenager. These developmental differences also impact the process by which technology is designed. The majority of the published methods for designing with and for children are for children in elementary or primary school (ages 6–11). Despite this emphasis in the literature there are children that fall beyond this more focused cross-section of children, including younger children and teenagers, who have been included in technology design processes.

Young children Regardless of how young children are, they can and should still be consulted in the design of their technology. When we refer to young children here, we are referring to children under six years of age; those who are in preschool or not yet traditional grade schools. Young children have unique needs as they do not have the same communication and developmental skills that slightly older children have. While this may require modifications to already intensive design processes, the unique needs of young children requires that their input be voiced in the design of technology. Raffle et al. (2011a) iteratively worked with young children to create prototypes of messaging systems and called for more research and design to include younger children in the design process. Guha et al. (2004) and Farber et al. (2002) provide practical guidelines for designing with these young children. Additionally, designers of technology for the youngest of children need to bear in mind the prevailing developmental theories, and design technology accordingly (Gelderblom, 2004; Gelderblom & Kotze, 2009). Developmental theorists such as Piaget (Piaget, 1973; Piaget, Gruber, & Vonèche, 1977) and Vygotsky (1978, 1986), are commonly referenced and considered when designing technology for young children.

Teenagers As young children require special consideration in technology design, so to do teenagers. Their continued development, coupled with typically busy schedules, make teenagers a unique group in the design of technology. Mazzone et al. (2008a) worked with teenagers in technology design and noted the unique needs of this group, including challenging schedules and need for attention to design methods and communication used with them. Teenagers also increasingly have access to ubiquitous technology, such as mobile phones, not only in conjunction with their families but as their own personal technology. Work has been done in participatory design with teenagers (Knudtzon et al., 2003; Iversen & Smith, 2012; Yip, Foss, & Guha, 2012), as well as workshops held on the topic of teenagers in design

(Read, Horton, Iversen, Fitton, & Little, 2013). Design with teenagers is a more nascent field (Yarosh et al., 2011). As with younger children, special attention must be paid to the development of teenagers in designing with and for them.

Family situations

Families are composed of diverse users but rather than designing for each distinct individual, a holistic approach can help address the collective needs of the family (Fails, 2013). Families are diverse in many ways including how they are composed, the developmental stages of the constituents, the time they have available, their culture and parenting style, as well as the kind of access they have to technology. This diversity creates additional challenges that need to be overcome in designing for families as a whole. In order to facilitate designing for families, often families with certain characteristics (e.g. single-parent families) are identified and involved in the design process or investigation of technologies. The most common are families with active children and a common task that has been explored is the sharing and managing of calendars (Neustaedter & Brush, 2006; Neustaedter, Brush, & Greenberg, 2007, 2009) as well as the sharing and managing of other types of family information and media (Brush & Inkpen, 2007; Egelman, Bernheim, & Inkpen, 2008). There is also research in more specialized scenarios such as divorced parents (Yarosh, 2008), children separated from parents (Yarosh, 2011b), how to stay connected with family from a distance via media spaces (Judge, Neustaedter, & Kurtz, 2010; Neustaedter & Judge, 2010), or even how to stay connected in extended family situations (Tee, Brush, & Inkpen, 2009). Families have also been identified as being important in the design of media spaces and interactions for this digital age (Takeuchi, 2011). For example, interfaces have been designed and developed to support coviewing experiences for children and parents both while they are copresent (Takeuchi & Stevens, 2011) and separated by space (Raffle et al., 2010, 2011b).

Using a corollary to Druin's classifications, Isola & Fails (2012) surveyed the literature from the Interaction Design and Children (IDC) conference and the Human Factors in Computing Systems (CHI) conference. They found that approximately 30% included families as users, 50% as testers, 10–15% as informants, and 5–10% as design partners. While there is a trend for growing interest in designing for families in some areas, there is a need for continued exploration of this important holistic user group (Isola & Fails, 2012). Yip et al. (2016) noted particular differences in the dynamics between parents and children and researchers and parents when codesigning with family members (Yip et al., 2016), noting that separating parents and children is beneficial at times, and that parents require researcher facilitation, and perhaps more time than children do to acclimate to the codesign environment.

Technology Evaluation Methods with Children

An area that is sometimes considered in conjunction with design methods is evaluation. The study of evaluation methods concerning technology for children in itself is a large field, with studies dedicated to reporting on evaluation methods (Sim & Horton, 2012; Zaman, 2009). As with technology design processes in general, children can and should be involved in evaluating their technology; however, modifications should be made to traditional evaluation methods designed to be used with adults in order for these evaluations to be valid and usable for children. For the purposes of this article, we will consider two primary types of evaluation: formative evaluation, which is used to give feedback iteratively on prototypes and technology throughout the design process; and summative evaluation, done when the design process is complete on a finished technology.

Formative evaluation is an integral part of the technology design process. Throughout the design process, and as an important part of the process, ideas and prototypes are continually evaluated, considered, and reconsidered through a variety of techniques. Techniques for formative evaluation, which occurs iteratively during the design process, include sticky noting (Fails et al., 2013), in which designers quickly jot evaluative ideas on small pieces of sticky paper, which can then be aggregated to show areas of need and possibility in the design, and line judging (Walsh et al., 2013) where designers use their bodies to physically order ideas or technologies along on a continuum. Formative evaluation in the iterative design process can help to avoid large mistakes and missteps in the technology. Without formative evaluation, technology can make it to the production stage with flaws that can and should have been caught during the design process.

Summative evaluation typically occurs at the end of the technology design process and gives designers feedback on a completed technology. Traditionally in the field of HCI, user studies have been conducted to evaluate technology, often employing methods such as observation. Many researchers have made modifications to classic usability tools so that they can be used appropriately with children. Read, MacFarlane, & Casey (2002) created a set of evaluative measures for use with children, called the Fun Toolkit. The tools include a Smiley-o-Meter, which is a Likert scale adapted for children; a Fun Sorter, in which children are asked to rank the relative fun of a variety of activities; and the Again-Again table, in which children are asked to tell if they would choose to do an activity again. The This-or-That method of evaluation (Zaman & Abeele, 2007) is a mixed-method model of evaluation with preschoolers, which includes a modified pairwise comparison tool in which young children are asked a series of questions to determine their preference between two technologies, as well as qualitative components including an interview and behavioral choice (Zaman, 2009). Other methods for summative evaluation with children include surveys, questionnaires, and diaries (Read & Markopoulos, 2011). Read and Markopolous (2008) also emphasize that while many of the basics of evaluation may stay the same over time, researchers should consider, update, and modify evaluation methods as technology progresses and children's worlds change over time. Recent updates include keeping evaluation sessions short, pairing children with a peer for evaluation sessions, and keeping evaluation techniques as simple as possible.

Current and Future Trajectories

In this chapter, we have thus far discussed technology design processes involving children through a historical lens and as the field exists today. In this section, we consider trends and future trajectories emerging in this area, including designing across distance, children as makers, formal and informal settings, and design thinking.

Designing together from afar

Technology design processes for children typically include a team of people, adults and sometimes children, working in a colocated manner. As the need to design for a more global audience grows, designers will need to consider how to include global voices in the design process. Partnering with children and adults from around the globe can encourage empathy and well-being (Guha et al., 2013). Efforts have already begun in this area, including the development of tools and systems that can support online, asynchronous collaboration (Walsh et al 2012; Walsh & Foss 2015), or facilitating user feedback while they use software in a distant location (Heintz et al., 2014).

Children as Makers

Currently, there is a growing movement of "makers." Making in the context of technology typically involves creating some kind of physical component, such as a robot or piece of clothing, which has a technologically embedded component that requires some knowledge of programming and/or circuitry. "Making" is a process that includes both creativity, physical components, and problem solving such as engineering and digital fabrication (Brahms, 2014). The maker movement can be appealing as it allows the maker to design his or her own technology, thereby creating an entirely personalized artifact. The lowering cost of tools such as electronic kits and 3D printers are allowing the maker movement to move into public spaces such as libraries, and even private spaces such as homes.

While makers are generally conceived of as adults, and perhaps even more specifically young adults, we believe that this trend of availability and usage will continue to expand to include children, thus including children in the design of their own, personal technology. There are high entry levels for 3D printing and other traditional making platforms like Arduino and they assume a certain level of knowledge and/or developmental level that many children do not have. This has prompted simpler kits to be developed with children in mind. Kits such as the LilyPad Arduino (Buechley, Eisenberg, Catchen, & Crockett, 2008) and i*CATch (Ngai et al., 2010), exist to support novice and young makers, especially in the e-textile arena by packaging smaller functioning components that lower the floor to entry into making. We believe the development of maker kits and making for young children is a trend that will continue, and maker tools will become more accessible to younger audiences. We have begun to see forays into this area, with kits such as with MakerShoe (Kazemitabaar, Norooz, Guha, & Froehlich, 2015), and Blocky Talky (Deitrick, Sanford, & Shapiro, 2014). The MakerShoe is a shoe conceived to be "hacked," manipulated, and redesigned by its young owners. Blocky Talky is a toolkit that supports the creation of digital and physical artifacts that can communicate wirelessly, thus lowering the floor to entry to allow novice users to design systems for the "Internet of Things." Using Blocky Talky, systems have been designed and developed that support applications varying from musical applications to programming a robot to feed non-colocated pets.

Informal and formal settings

In most of the design methods that we have addressed in this chapter, children are involved in a design process that requires more than simply answering questions and giving feedback. Many of the design methods we have described are used in different settings such as formal and informal educational settings. For example Cooperative Inquiry is typically employed in an informal setting that could be classified as an after-school program or activity; however, forays have also been made in taking cooperative inquiry into more formal school settings (Foss et al., 2013; Guha et al., 2004). There are many design and technology after-school programs such as RoboFun (www.robofun.org) and Computer Clubhouse (www.computerclubhouse.org), which are informal yet provide educational experiences where students learn design principles, problem-solving skills, and create systems and artifacts.

Design thinking

Decades ago, Cross (1982) called for design to be considered a major component of education. In our work, we have suggested moving design partnering with children into both informal and formal education systems (Guha, Druin, & Fails, 2011). Children who are more integrally involved throughout a design process have experience with many more cognitive and social process, including problem solving and creativity (Guha, 2010). Taken together, the experiences of being on a technology design team expose children to *design thinking*, which involves exploring a variety of solutions through iterations, and encouraging children to become real agents of change (Carrol et al., 2010). Today there is a rekindled emphasis on making and doing. Schools in the United States focus heavily on STEM (science, technology, engineering, and math). Though in past years engineering may have been overlooked, currently it is covered in many schools, and has a design component. Careers today, and of the future, will require many complex skills and workers will need to design and conceive new and imaginative solutions to the challenges they will face. This is the kind of design thinking that children are exposed to as part of a technology design team (Guha et al., 2017).

Design thinking is a notion whose time has come. We expect that children will become more involved in technology design not only because it produces innovative technology but also because it can serve as training and experience for children in a way of thinking and knowing that will serve them well throughout their lives. These principles of designing and creating will continue to be spread, shared, and cultivated with children both in formal and informal educational settings.

Conclusion

In this chapter, we have outlined the scope of participatory technology design processes for children. In so doing we have covered the history, current state, and future projections for technology design processes with children, as well as iterative technology evaluation methods for working with children. We presented the primary current participatory methods of design for children as well as some of the adaptations that are necessary for working with children with special circumstances, including children with special needs, of varying ages, and as part of a family. Future design methods for children will continue to accommodate formal and informal space, with increased emphasis on the settings such as home and play. Due to the increasing ease of communicating and the necessity of working with people around the world, this trend will also impact how children are involved in the design of technologies for them, meaning that they will also need to have platforms that support children as they collaborate across distances to design new and better technologies. As children create and design innovative solutions they will learn skills that will allow them to make and think creatively to address and overcome challenges they will face in the workplace, in school, and at home.

References

AlGhamdi, M., Beheshti, J., Cole, C., Abuhimed, D., & Lamoureux, I. (2013). Bonded design: Children as co-designers of an educational tool for students involved in an inquiry-based project. In Andrew Grove (Ed.), *Proceedings of the ASIS&T Annual Meeting: Beyond the Cloud: Rethinking Information Boundaries (ASIST '13)* (Article no. 118). Silver Springs, MD: American Society for Information Science.

Antle, A. (2003). Case study: The design of CBC4Kids' StoryBuilder. In S. MacFarlane, T. Nicol, J. Read & L. Snape (Eds.), *Proceedings of the Conference on Interaction Design and Children (IDC '03)* (pp. 59–68). New York, NY: ACM.

Antle, A. (2006). Child-personas: Fact or fiction? *Proceedings of Designing Interactive Systems (DIS 2006)* (University Park, PA, USA) (pp. 22–30). New York, NY: ACM.

Antle, A. (2008). Child-based personas: Need, ability, and experience. *Cognition, Technology, and Work* [Special issue on child-computer interaction: methodological research], *10*(2), 155–166.

Battacharya, A., Gelsomini, M., Perez-Fuster, P., Abowd, G., & Rogza, A. (2015). Designing motion-based activities to engage students with autism in classroom settings. In *Proceedings of the Conference on Interaction Design and Children (IDC 2015)* (Boston, MA) (pp. 69–78). New York, NY: ACM.

Beheshti, J., & Large, A. (2011). An interface for opportunistic discovery of information for young people. In Julie A. Jacko (Ed.), *Proceedings of the 14th International Conference on Human-Computer Interaction: Users and Applications* (pp. 541–548). Berlin: Springer-Verlag.

Benton, L., Johnson, H., Ashwin, E., Brosnan, M., & Grawemeyer, B. (2012). Developing IDEAS: Supporting children with autism within a participatory design team. In *Conference on Human Factors in Computing Systems (CHI)* (Austin, TX) (pp. 2599–2608). New York: ACM.

Benton, L., Johnson, H., Brosnan, M., Ashwin, E., & Grawemeyer, B. (2011). IDEAS: An interface design experience for the autistic spectrum. In *Conference Extended Abstracts on Human Factors in Computing Systems (CHI-EA)* (Vancouver) (pp. 1759–1764). New York: ACM.

Bjerknes, G., Ehn, P., & Kyung, M. (1987). *Computers and democracy: A Scandinavian challenge*. Aldershot: Alebury.

Bødker, S., Ehn, P., Sjögren, D., & Sundblad, Y. (2000). *Co-operative design—Perspectives on 20 years with "the Scandinavian IT Design Model."* Paper presented at the Nordic Forum for Human-Computer Interaction (NordiCHI 2000), Stockholm.

Borjesson, P., Barendregt, W., Eriksson, E., & Torgersson, O. (2015). Designing technology for and with developmentally diverse children—A systematic literature review. In *Proceedings of the Conference on Interaction Design and Children (IDC 2015)* (Boston, MA) (pp. 79–88). New York, NY: ACM.

Brahms, L. (2014). *Making as a learning process: Identifying and supporting family learning in informal settings* (Unpublished doctoral dissertation). University of Pittsburgh, Pittsburgh, PA.

Brush, A. J., & Inkpen, K. (2007). Yours, mine and ours? Sharing and use of technology in domestic environments. In *Proceedings of the ACM Conference on Ubiquitous Computing (UbiComp 2007)* (pp. 109–126). Berlin: Springer-Verlag.

Buechley, L., Eisenberg, M., Catchen, J., & Crockett, A. (2008). The LilyPad Arduino: Using computational textiles to investigate engagement, aesthetics, and diversity in computer science education. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2008)* (Florence, Italy) (pp. 423–432). New York, NY: ACM.

Carroll, M., Goldman, S., Britos, L., Koh, J., Royalty, A., & Hornstein, M. (2010). Destination, imagination and the fires within: Design thinking in a middle school classroom. *The International Journal of Art and Design Education, 29*(1), 37–53.

Cramer, M. D., Hirano, S., Tentori, M. Yeganayan, M., & Hayes, G. R. (2011). Classroom-based assistive technology: Collective use of interactive visual schedules by students with autism. In *Proceedings of ACM Conference on Human Factors in Computing (CHI 2011)* (Vancouver, B.C., Canada) (pp. 1–10). New York, NY: ACM.

Cross, N. (1982). Designerly ways of knowing. *Design Studies, 3*(4), 221–227.

Deitrick, E., Sanford, J., & Shapiro, R. B. (2014, June). *BlockyTalky: A low-cost, extensible, open source, programmable, networked toolkit for tangible creation.* Interactive demonstration at ACM Interaction Design and Children (IDC 2014), Aarhus, Denmark.

Druin, A. (1999). Cooperative inquiry: developing new technologies for children with children. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (CHI 1999; Pittsburgh, PA) (pp. 592–599). New York, NY: APA.

Druin, A. (2002). The role of children in the design of new technology. *Behaviour and Information Technology, 21*(1), 1–25.

Druin, A., Bederson, B., Rose, A., & Weeks, A. (2009). From New Zealand to Mongolia: Co-designing and employing a digital library for the world's children. *Children, Youth, and Environment: Special Issue on Children in Technological Environments, 19*(1), 34–57.

Egelman, S., Bernheim, A. J., Inkpen, K. (2008). Family accounts: A new paradigm for user accounts within the home environment. *Proceedings of ACM Computer Supported Cooperative Work (CSCW 2008)* (San Diego, CA) (pp. 669–678). New York, NY: ACM.

Fails, J. A. (2013, April). *Opportunities and challenges in designing technologies for families.* Paper presented at Exploring the Diversity of Families: Designing Technologies for the Contemporary Family Life Workshop in coordination with the Conference on Human Factors in Computing Systems (CHI), Paris, France.

Fails, J. A., Guha, M. L., & Druin, A. (2013). Methods and techniques for involving children in the design of new technology for children. *Foundations and Trends in Human-Computer Interaction*, *6*(2). Retrieved from http://dx.doi.org/ 10.1561/1100000018

Farber, A., Druin, A., Chipman, G., Julian, D., & Somashekhar, S. (2002). How young can our design partners be? In T. Binder, J. Gregory, I. Wagner (Eds.) *Proceedings of the Participatory Design Conference*, Malmo, Sweden. Palo Alto, CA: CPSR.

Foss, E., Guha, M. L., Papadatos, P., Clegg, T., Yip, J., & Walsh (2013). Cooperative inquiry extended: Creating technology with middle school students with learning differences. *Journal of Special Education Technology, 28*(3), 33–46.

Frauenberger, C., Good, J. and Alcorn, A. (2012). Challenges, opportunities and future perspectives in including children with disabilities in the design of interactive technology. *Proceedings of the Conference on Interaction Design and Children (IDC 2012)* (Bremen, Germany) (pp. 367–370). New York, NY: ACM.

Frauenberger, C., Good, J., Alcorn, A. and Pain, H. (2012). Supporting the design contributions of children with autism spectrum conditions. *Proceedings of the Conference on Interaction Design and Children (IDC 2012)* (Bremen, Germany) (pp. 134–143). New York, NY: ACM.

Gelderblom, H. (2004). Designing software for young children: Theoretically grounded guidelines. *Proceedings of the Conference on Interaction Design and Children (IDC 2004)* (College Park, MD) (pp. 121–122). New York, NY: ACM.

Gelderblom, H., & Kotze, P. (2009). Ten design lessons from the literature on child development and children's use of technology. *Proceedings of the Conference on Interaction Design and Children (IDC 2009)* (Como, Italy) (pp. 52–60). New York, NY: ACM.

Gibson, F. (2007). Conducting focus groups with children and young people: Strategies for success. *Journal of Research in Nursing, 12*(5), 473–483.

Guha, M. L., Druin, A., Chipman, G., Fails, J. A., Simms, S., & Farber, A. (2004). Mixing ideas: A new technique for working with young children as design partners. *Proceedings of the Conference on Interaction Design and Children (IDC 2004)* (College Park, MD) (pp. 35–42). New York, NY: ACM.

Guha, M. L. (2010). *Understanding the social and cognitive experiences of children involved in technology design processes* (Doctoral dissertation), University of Maryland, College Park, MD.

Guha, M. L., Druin, A., & Fails, J.A. (2011). How children can design the future. In J. Jacko (Ed.). *Human-Computer Interaction, Part IV, International Conference on Human-Computer Interaction* (pp. 559–569). Berlin, Springer.

Guha, M. L., Druin, A., & Fails, J.A. (2013). Cooperative inquiry revisited: Reflections of the past and guidelines for the future of intergenerational co-design. *International Journal of Child-Computer Interaction, 1*(1), 14–23.

Guha, M. L., McNally, B., and Fails, J. A. (2017). "Design Partners in Schools: Encouraging Design Thinking Through Cooperative Inquiry" in Goldman, S. and Kabayadondo, Z. (2017). *Taking Design Thinking to School: How the Technology of Design Can Transform Teachers, Learners, and Classrooms.* New York, NY: Routledge.

Gutman, B. (2012). *Fisher-Price on connecting with moms in the digital world.* Retrieved from http://www.forbes.com/sites/marketshare/2012/09/13/fisher-price-on-connecting-with-moms-in-the-digital-world/#18139b18705f

Heintz, M., Law, E. L. C., Govaerts, S., Holzer, A., & Gillet, D. (2014). Pdot: participatory design online tool. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI 2014)* (Toronto, Canada) (pp. 2581–2586). New York, NY: ACM.

Hourcade, J. P., Bullock-Rest, N. E., & Hansen, T. E. (2012). Multitouch tablet applications and activities to enhance the social skills of children with autism spectrum disorders. *Personal Ubiquitous Computing, 16*(2), 157–168.

Isola, S., & Fails, J. (2012). Family and design in the IDC and CHI communities. *Proceedings of Interaction Design and Children (IDC)* (Bremen, Germany) (pp. 40–49). New York, NY: ACM.

Iversen, O. S. and Smith, R. C. (2012). Scandinavian participatory design: dialogic curation with teenagers. *Proceedings of the Conference on Interaction Design and Children (IDC 2012)* (Bremen, Germany) (pp. 106–115). New York, NY: ACM.

Judge, T. K., Neustaedter, C., and Kurtz, A. (2010). The family window: The design and evaluation of a domestic media space. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2010)* (Atlanta, GA) (pp. 2361–2370). New York, NY: ACM.

Kazemitabaar, K., Norooz, L., Guha, M. L., & Froehlich, J. (2015). MakerShoe: Toward a wearable e-textile construction kit to support creativity, playful making, and self-expression [Demonstration paper]. In *Proceedings of the 14th International Conference on Interaction Design and Children* (Boston, MA) (pp. 449–452). New York, NY: ACM.

Kientz, J. A. (2012). Emebedded capture and access: Encouraging recording and reviewing of data in the caregiving domain. *Personal Ubiquitous Computing, 16*(2), 209–221.

Kientz, J. A., Goodwin, S., Hayes, G. R., & Abowd, G. (2013). *Interactive technologies for autism* (1st ed.). San Rafael, CA: Morgan & Claypool.

Knudtzon, K., Druin, A., Kaplan, N., Summers, K., Chisik, Y., Kulkarni, R., …. Bederson, B. (2003). Starting an intergenerational technology design team: A case study. *Proceedings of the Conference on Interaction Design and Children (IDC 2003)* (Preston, England) (pp. 51–58). New York, NY: ACM.

Large, A., Bowler, L., Beheshti, J. and Nesset, V. (2007). Creating web portals with children as designers: Bonded design and the zone of proximal development. *McGill Journal of Education, 42*, 61–82.

Large, A., Nesset, V., Beheshti, J., & Bowler, L. (2006). "Bonded Design": A novel approach to intergenerational information technology design. *Library and Information Science Research, 28*, 64–82.

Luchini, K., Quintana, C., & Soloway, E. (2004). Design guidelines for learner-centered hand-held tools. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2004)* (Vienna, Austria) (pp. 135–142). New York, NY: ACM.

Malone, T. W. (1982). Heuristics for designing enjoyable user interfaces. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '82)* (pp. 63–68). New York, NY: ACM.

Mazzone, E., Read, J., & Beale, R. (2008a). Design with and for disaffected teenagers. In *Proceedings of the 5th Nordic Conference on Human-Computer Interaction: Building Bridges* (pp. 290–297). New York, NY: ACM.

Mazzone, E., Read, J., & Beale, R. (2008b). Understanding children's contributions during informant design. In *Proceedings of the 22nd British HCI Group Annual Conference on People and Computers: Culture, Creativity, Interaction—Volume 2 (BCS-HCI '08)* (pp. 61–64). Swindon: BCS Learning and Development Ltd.

Moser, C., Fuchsberger, V., & Tscheligi, M. (2011). Using probes to create child personas for games. In *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology (ACE '11)* (pp. 39:1–39:8). New York, NY: ACM.

Muller, M. J., & Druin, A. (2012). Participatory design: The third space in HCI. In J. Jacko & A. Sears (Eds.), *The human-computer interaction handbook* (pp. 1050–1075). Hillsdale, NJ: Lawrence Erlbaum.

Naranjo-Bock, C. (2011). *Approaches to user research when designing for children*. Retrieved from www.uxmatters.com/mt/archives/2011/03/approaches-to-user-research-when-designing-for-children.php

Neustaedter, C., & Brush, A. J. (2006). "LINC-ing" the family: the participatory design of an inkable family calendar. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2006)* (Montréal, Québec) (pp. 141–150). New York, NY: ACM.

Neustaedter, C., Brush, A. J., & Greenberg, S. (2007). A digital family calendar in the home: lessons from field trials of LINC. In *Proceedings of the ACM Conference on Graphics Interface (GI 2007)* (Montréal, Québec) (pp. 199–206). New York, NY: ACM.

Neustaedter, C., Brush, A. J., & Greenberg, S. (2009). The calendar is crucial: Coordination and awareness through the family calendar. *ACM Transactions on Human-Computer Interaction, 16*(1), Article no. 6.

Neustaedter, C., & Judge, T. K. (2010) Peek-a-boo: The design of a mobile family media space video. In *Proceedings of the ACM Conference on Ubiquitous Computing (UbiComp 2010)* (Copenhagen, Denmark). doi:10.1145/1864431.1864482

Ngai, G., Chan, S. C. F., Ng, V. T. Y., Cheung, J. C. Y., Choy, S. S. S., Lau, W. W. Y., & Tse, J. T. P. (2010). i*CATch: a scalable plug-n-play wearable computing framework for novices and children. In *Proceedings of the Conference on Human Factors in Computing (CHI 2010)* (Atlanta, GA) (pp. 443–452). New York, NY: ACM.

Norooz, L., Mauriello, M., Jorgenson, A., McNally, B., & Froehlich, J. (2015). BodyVis: A new approach to body learning through wearable sensing and visualization. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2015)* (Seoul, South Korea) (pp. 1025–1034). New York, NY: ACM.

Piaget, J. (1973). *To understand is to invent: The future of education*. New York, NY: Grossman.

Piaget, J., Gruber, H. E., & Vonèche, J. J. (1977). *The essential Piaget*. New York, NY: Basic Books.

Raffle, H., Ballagas, R., Revelle, G., Horii, H., Follmer, S., Go, J.,… Spasojevic, M. (2010). Family story play: Reading with young children (and elmo) over a distance. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2010)* (Atlanta, GA) (pp. 1583–1592). New York, NY: ACM.

Raffle, H., Ballagas, R., Revelle, G., Mori, K., Horii, H., Paretti, C., & Spasojevic, M. (2011a). Pop goes the cell phone: Asynchronous messaging for preschoolers. In *Proceedings of the Conference on Interaction Design and Children (IDC '11)* (Ann Arbor, MI) (pp. 99–108). New York, NY: ACM.

Raffle, H., Revelle, G., Mori, K., Ballagas, R., Buza, K., Horii, H.,… Spasojevic, M. (2011b). Hello, is grandma there? Let's read! StoryVisit: Family video chat and connected e-books. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2011)* (Vancouver, BC) (pp. 1195–1204). New York, NY: ACM.

Read, J., Horton, M., Iversen, O. S., Fitton, D., & Little, L. (2013) Methods of working with teenagers in interaction design (Workshop). *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2013)* (Paris, France) (pp. 3243–3246). New York, NY: ACM.

Read, J., MacFarlane, S., & Casey, C. (2002). Endurability, engagement, and expectations. In *Proceedings of Interaction Design and Children (IDC 2002)* (Eindhoven, The Netherlands) (pp. 1–23). Maastricht: Shaker Publishing.

Read, J. C. and Markopoulos, P. (2008). Lifelong interactions: Understanding children's interactions: Evaluating children's interactive products. *Interactions, 15*(6), 26–29.

Read, J. C., & Markopoulos, P. (2011). Evaluating children's interactive products [Course notes]. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2011)* (Vancouver, Canada) (n.p.). New York, NY: ACM.

Rideout, V. (2013). *Zero to eight: Children's media use in America 2013. Common sense media*. Retrieved from www.commonsensemedia.org/research/zero-to-eight-childrens-media-use-in-america-2013

Scaife, M., & Rogers, Y. (1999). Kids as informants: Telling us what we didn't know or confirming what we knew already? In A. Druin (Ed.), *The design of children's technology* (pp. 27—50). San Francisco, CA: Morgan Kauffman Publishers.

Scaife, M., Rogers, Y., Aldrich, F., & Davies, M. (1997). Designing for or designing with? Informant design for interactive learning environments. *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 97)* (Atlanta, Georgia) (pp. 343–350). New York, NY: ACM.

Schuler, D., & Namioka, A. (Eds.). (1993). *Participatory design: Principles and practices.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Sim, G., & Horton, M. (2012). Investigating children's opinions of games: Fun Toolkit vs. This or That. *Proceedings of the Conference on Interaction Design and Children (IDC 2012)* (Bremen, Germany) (pp. 70–77). New York, NY: ACM.

Smith, A. (2011). *Americans and their cell phones* (Pew Internet research report). Retrieved from http://www.pewinternet.org/2011/08/15/americans-and-their-cell-phones/

Smith, R., Iversen, O., & Lynggaard, A. (2013). Towards an ecological inquiry in child-computer interaction. In *Proceedings of the Conference on Interaction Design and Children (IDC 2013)* (New York City, NY) (pp. 183–192). New York, NY: ACM.

Soloway, E., Guzdial, M. and Hay, K. E. (April 1994). Learner-centered design: The challenge for HCI in the 21st century. *Interactions, 1*(2), 36–48.

Soloway, E., Jackson, S. L., Klein, J., Quintana, C., Reed, J., Spitulnik, J., … Scala, N. (1996). Learning theory in practice: Case studies of learner-centered design. In *Proceedings of the ACM Conference on Human Factors in Computing Systems: Common Ground (CHI 1996)* (Vancouver) (pp. 189–196). New York, NY: ACM.

Takeuchi, L. (2011, May). *Families matter: Designing media for a digital age*. Joan Ganz Cooney Center at Sesame workshop. Retrieved from http://www.joanganzcooneycenter.org/publication/families-matter-designing-media-for-a-digital-age/

Takeuchi, L., & Stevens, R. (2011, December). The new coviewing: Designing for learning through joint edia Engagement. Joan Ganz Cooney Center at Sesame Workshop. Retrieved fromhttp://www.joanganzcooneycenter.org/publication/the-new-coviewing-designing-for-learning-through-joint-media-engagement/

Tee, K., Brush, A. J., & Inkpen, K. (2009). Exploring communication and sharing between extended families. *International Journal of Human-Computer Studies, 67*(2), 128–138.

Vygotsky, L. S. (1978). *Mind in society.* Cambridge, MA: Harvard University Press.

Vygotsky, L. S. (1986). *Thought and language.* Cambridge, MA, MIT Press.

Walsh, G., Druin, A., Guha, M. L., Bonsignore, E., Foss, E., Yip, J., … Brewer, R. (2012). DisCo: A co-design tool for online, asynchronous distributed child and adult design partners. In *Proceedings of the Conference on Interaction Design and Children (IDC 2012)* (Bremen, Germany) (pp. 11–19). New York, NY: ACM.

Walsh, G., Druin, A., Guha, M. L., Foss, E., Golub, E., Hatley, L., … Franckel, S. (2010). Layered elaboration: A new technique for co-design with children. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI 2010)* (Atlanta, GA) (pp. 1237–1240). New York, NY: ACM.

Walsh, G., & Foss, E. (2015). A case for intergenerational distributed co-design: The online Kidsteam example. *Proceedings of the Conference on Interaction Design and Children (IDC 2015)* (Boston, MA) (pp. 99–108). New York, NY: ACM.

Walsh, G., Foss, E., Yip, J., & Druin, A. (2013). FACIT PD: Framework for analysis and creation of intergenerational techniques for participatory design. In *Proceedings of Conference on Human Factors in Computing Systems (CHI 2013)* (Paris) (pp. 2893–2902). New York, NY: ACM.

Wärnestål, P., Svedberg, P., & Nygren, J. (2014). Co-constructing child personas for health-promoting services with vulnerable children. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems (CHI 2014)* (Toronto, Canada) (pp. 3767–3776). New York, NY: ACM.

Xu, D., Mazzone, E., & MacFarlane, S. (2005). Informant design with children—designing children's tangible technology. In *Proceedings of the First International Workshop "Re-Thinking Technology in Museums"* (Limerick, Ireland) (pp. 141–146). Limerick: Ireland University of Limerick.

Yarosh, S. (2008). Supporting long-distance parent-child interaction in divorced families. In *Proceedings of Conference on Human Factors in Computing Systems (CHI EA)* (Florence, Italy) (pp. 3795–3800). New York, NY: ACM.

Yarosh, S., & Abowd, G. D. (2011). Mediated parent-child contact in work-separated families. In *Proceedings of Conference on Human Factors in Computing Systems (CHI 2011)* (Vancouver, BC, Canada) (pp. 1185–1194). New York, NY: ACM.

Yarosh, S., Radu, I., Hunter, S., & Rosenbaum, E. (2011). Examining values: an analysis of nine years of IDC research. In *Proceedings of Conference on Interaction Design and Children (IDC 2011)* (Ann Arbor, MI) (pp. 136–144). New York, NY: ACM.

Yip, J. C., Clegg, T., Ahn, J., Uchidiuno, J. O., Bonsignore, E., Beck, A., … Mills, K. (2016). The evolution of engagements and social bonds during child-parent co-design. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI 2016)* (San Jose, CA) (pp. 3607–3619). New York, NY: APA.

Yip, J. C., Foss, E., & Guha, M. L. (2012, October). *Co-designing with adolescents.* Paper presented at the Designing Interactive Technology for Teens Workshop, NordiCHI, Copenhagen, Denmark.

Zaman, B. (2009). Introduction and validation of a pairwise comparison scale for UX evaluations and benchmarking with preschoolers. In T. Gross (Ed.) *Interact* (pp. 634–637). Uppsala: Springer.

Zaman, B., & Abeele, V. (2007). How to measure the likeability of tangible interaction with preschoolers. In *Proceedings of CHI Nederland* (pp. 57–59). Amsterdam: Infotec Nederland.

Part III
Evaluation Factors

10

User Experience

Jakob Grue Simonsen

Two Stories on User Experience

Before tackling the general problem of understanding and demarcating the notion of User Experience, it is fruitful to consider two stories that describe the difference between User Experience and the more traditional notion of usability.

Story A

P is the owner of, and avid user of, several smartphones, a tablet, a laptop, and a desktop computer at her home. She is a serious ESports enthusiast and player, using her desktop computer to participate in highly competitive online games, and she and her team travel to regional and national contests with her desktop setup to compete for prizes.

For P, the usability of her desktop PC hardware is of greater importance than for both the casual leisure user and most office professionals using their computers for work purposes: to be competitive in esports, she must maintain a high amount of very precise mouse clicks and keystrokes per second, her Internet connection must be completely lag free, and her screen and graphics hardware must maintain a very high, constant frame rate. However, her use of the PC, its hardware, the game software she competes with, and ancillary software, are all means to other ends: her personal satisfaction and reward comes from overcoming steeper challenges, failing against stronger opponents, but subsequently mastering harder techniques to improve her abilities. She revels in hours-long protracted matches that contain dozens of focused, short-duration challenges that require her complete attention, blisteringly fast conscious and unconscious decision making, and physical skill. The ergonomic finish of her hardware is paramount, but P values the sleek design of her keyboard and mouse, both for her own pleasure when spending hours at home honing her competitive skill, and for showing off at contests. For P, the experience of her interaction with the PC as a technological artifact encompasses much more than mere usability—it contains social interactions with her team mates and opponents, strong emotional responses related to frustration, loss, and elation, pride in her abilities, intense physical and intellectual effort, and appreciation of the aesthetics of good industrial design.

The Wiley Handbook of Human Computer Interaction, Volume 1, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

Story B

R is retired and owns few electronic devices bar a radio, a television set, and a smartphone given to him by his daughter. An avid hobby angler, R uses his phone for calling friends, children, and grandchildren, and uses a few specialized apps for checking local weather and angling conditions. In the last year, R's grandchildren, all living several hundred miles away, have taken to sending him pictures and texts from special occasions, and share some personal moments including school and extracurricular accomplishments. R takes great interest in answering these using phone calls, pictures, text, and occasionally videos, but is completely uninterested in online social media. For R, the ease of use and effectiveness with which he can contact, and be contacted by, his grandkids, and use the phone to support his hobbies, is important, but the total experience involves strong interpersonal bonds, emotions towards family members, and self-actualization in pursuing his interests. The experience of using the smartphone is not merely a consequence of his interaction with the smartphone as a technological artifact, but rather a consequence of his joy and fulfillment in maintaining personal relationships, and the heightened sense of closeness and presence that he experiences when physically separated from his family.

The many definitions of user experience

Juxtaposition of "user" and "experience" in connection with design and production of technological artifacts has occurred at least as early as the beginning of the 20th century; descriptions of integration of user satisfaction and experience as part of the product design process appear even earlier in the study of ergonomics.

In digital computing, similar uses of "user experience" and "user's experience" appear sporadically in the academic literature in the 1980s and early 1990s—for example, Laurel (1986) and Whiteside and Wixon (1987). The modern impetus for academic study of "User Experience" as a concept appears to be a paper by Norman, Miller, and Henderson Jr. (1995, p. 155) at CHI '95 as part of a description of the design practice at Apple Computer Inc. at the time ("critical aspects of human interface research and application at Apple or, as we prefer to call it, the 'User Experience'"); it appears that the time was in internal use at Apple at least as early as 1993.

After the popularization of the term "User Experience" in the mid- to late 1990s, a number of academic papers set out to have the research community take the concept seriously, and pushed for serious research agendas—see, for example, Alben (1996) and Overbeeke, Djajadiningrat, Hummels, and Wensveen (2000), and the early years of the new millennium witnessed a number of efforts to ground and define the concept, for example Forlizzi and Battarbee (2004), Hassenzahl (2004), and Wright, McCarthy, and Meekison (2004). The 2010s have seen even further attempts at demarcating the notion of User Experience but there is still no commonly agreed definition or scoping of the concept.

The ISO standard on ergonomics of human-system interaction (ISO 9241-210, p. 3 (ISO standard)) defines user experience, in its revision of 2009, as "a person's perceptions and responses that result from the use or anticipated use of a product, system or service." The ISO definition is sufficiently vague that it can be interpreted in a variety of ways; however, it is instructive to contrast it with the ISO 9241 definition of *usability*—"the effectiveness, efficiency and satisfaction with which specified users achieve specified goals in particular environments, p. 3 (ISO standard)." While *satisfaction* can certainly be a "response" in the sense of User Experience, the usability definitions of "effectiveness" and "efficiency" are clearly more specific, and narrower, terms emphasizing task completion and efficiency rather than "a person's perceptions and responses" in the definition of User Experience; the differences between the two ISO definitions mimic one of the seemingly few almost-universal consensuses in User Experience research, namely that User Experience is not usability (but may have overlapping concerns), amusingly expressed by Hassenzahl and Roto (2007, p. 12) as: "Usability wants us to die rich. UX wants us to die happy." See Chapter 3 for more information about the ISO standards.

The academic literature is replete with definitions, or at least descriptions with the semblance of definitions, from the mid-1990s until about 2010. For example, an early definition by Alben (1996, p. 12) from the mid-1990s read:

> All the aspects of how people use an interactive product: the way it feels in their hands, how well they understand how it works, how they feel about it while they're using it, how well it serves their purposes, and how well it fits into the entire context in which they are using it.

Similarly, McNamara, and Kirakowski (2006, p. 26) wrote that "The user experience considers the wider relationship between the product and the user in order to investigate the individual's personal experience of using it." In the same year, Hassenzahl and Tractinsky (2006, p. 95) described User Experience as

> A consequence of the user's internal state (predispositions, expectations, needs, motivation, mood, and so on), the characteristics of the designed system (for example, complexity, purpose, usability, functionality, and so on), and the context (or the environment) within which the interaction occurs (for example, organizational / social setting, meaningfulness of the activity, voluntariness of use, and so on).

And Hassenzahl (2008b) contended that

> UX [is] a momentary, primarily evaluative feeling (good-bad) while interacting with a product or service…Good UX is the consequence of fulfilling the human needs for autonomy, competency, stimulation (self-oriented), relatedness, and popularity (others-oriented) through interacting with the product or service (i.e., hedonic quality). Pragmatic quality facilitates the potential fulfillment of be-goals.

From 2010 onwards, many authors seem to have accepted that a single, *short* definition that is precise and encompasses all aspects that the research and practitioners' community associate with "User Experience," is not forthcoming, and indeed does not appear possible. Thus, definitions appear that are deliberately fuzzy but acknowledge that User Experience is a multifaceted phenomenon with contributions from a variety of fields. For example, Wiklund-Engblom, Hassenzahl, Bengs, and Sperring (2009, p. 666) wrote that

> there is a common understanding that user experience is *holistic*—it emphasizes the totality of emotion, motivation, and action in a given physical and social context—and that it is *subjective*—focusing on the "felt experiences" (McCarthy & Wright, 2007) rather than product attributes…

and Robert and Lesage (2011b, p. 311) defined User Experience as "a multidimensional construct that defines the overall effect over time on the user of interacting with a system and service in a specific context."

Law and Van Schaik (2010, p. 313) held that "UX manifests as quality in design, in interaction, and in value, with diverse measures from many methods and instruments," but firmly state that a more thorough *model building* is needed for User Experience, both as regards measurement models (measuring the constructs of UX) and structural models (elucidating the causal relationship between constructs). Similar, but broader, concerns about the need for more thorough theory building in User Experience were raised by Obrist et al. (2012a, 2012b).

Several efforts have been made to chart the proliferation of views on, and definitions of, User Experience. For example, Law, Roto, and Väänänen-Vaino-Mattila (2009, p. 726) surveyed 270 User Experience researchers and practitioners, finding that respondents had not only highly varied personal accounts of what User Experience is, and is about, but also had highly varied criteria for what a sensible definition should contain (e.g., some respondents favored that a definition "should address what User Experience is rather than what causes user experience"). Particular points of disagreement pertained to the temporal aspect of User Experience, e.g., should it include experience prior to using an artifact? Experience (long) after such use? Should it include reflection on use?

Obrist et al. (2012b) found that 70 participants in a User Experience special interest groups session at the CHI '11 conference when asked to answer the question "What theoretical roots do we build on, if any in User Experience research?" provided (after coding) 56 theoretical perspectives of disparate scope, broadness, and scientific provenance (e.g., "actor-network theory;" "critical theory," and "cognitive load theory"), belonging to nine traditional scientific disciplines (anthropology, art, communication, design, education, marketing, philosophy, psychology, and sociology).

Quite apart from the discussion among academics and serious practitioners, the term "User Experience" and its abbreviation UX is bantered around in a large number of texts on practical user-centered design for various domains (often web pages or applications for mobile devices) where the term is used occasionally as synonymous with usability, or with user-centered design itself. This is by no means an erroneous approach in the sense that what is treated is almost invariably an aspect, or dimension, of User Experience according to all definitions above; but these texts very rarely clarify what their particular notion of "User Experience" is, if any.

Finally, while the reader may feel skeptical due to the many different accounts and definitions of User Experience, the following vindicating fact should be kept in mind: if a taxonomy of concepts in human-computer interaction were ever devised, User Experience will have a taxonomic rank *above* many other hotly debated topics whose definition has stabilized more, but which could also arguably be considered subfields of User Experience, for example usability; it is natural that a very general concept that properly contains many other important constructs has a necessarily broad definition, and that proper model construction, for example the structural models sought by Law and Van Schaik (2010) are correspondingly more difficult to devise. It seems vastly more fruitful to cease bemoaning the lack of a precise, brief definition of User Experience, and instead focus on giving precise definitions of its subfields or the aspects and dimensions that are its constituent parts.

Dimensions of User Experience

The historical lack of agreement on a single definition of User Experience has motivated a number of researchers to consider different aspects or dimensions of the term, and has led some to categorize these dimensions. A common understanding is that User Experience contains a number of dimensions not found in, say, usability, and there seems to be some consensus in the literature that "experiential" and "affective" dimensions are part of User Experience, but the connotations of these terms tend to differ from study to study. Another concern is that while some nomenclature seems well established within the HCI community studying User Experience, it may differ subtly from the nomenclature in the ancillary disciplines used for studying it: for example, "affect" and "emotion" appear to be treated as synonyms in much of the HCI literature on User Experience, but denote quite different phenomena in some psychology research (Scherer, 2005).

While most authors categorizing dimensions of User Experience focus on dimensions such as emotion and aesthetics, others define dimensions that pertain to the physiological, cognitive, or social state of the user in interaction with technology (Robert & Lesage, 2011a). Table 10.1 outlines some of the extant categorizations of dimensions of User Experience.

Bargas-Avila and Hornbæk (2011) note that a large number of studies propose new dimensions that are either variations of already known dimensions, or whose relationship to existing dimensions is not properly investigated. As one example, they note that the dimension of "enchantment" introduced by McCarthy, Wright, Wallace, and Dearden (2006), and used in several later studies, e.g. (Ross, Overbeeke, Wensveen, & Hummels, 2008; Ní Conchúir & McCarthy, 2008) has no well-defined relation to other constructs in User Experience, and no clear distinction from established concepts such as experience of flow (Czikszentmihalyi, 1997).

Different dimensions may have different impact on product design and quality; for example, using a broad distinction between pragmatic and hedonic dimensions, Hassenzahl, Schobel, and Trautmann (2008) found that study participants rated a hedonic design higher than a pragmatic design when participants were subjected to a promotion focus (concern for personal growth and attaining positive outcomes), with a reverse effect when participants were subjected to a prevention focus (concern for safety and avoidance of negative outcomes).

As expected, many of the dimensions of User Experience that have been studied empirically appear to be either orthogonal, or at best have tenuous connection to, traditional notions of usability. For example, a number of studies have considered the effect of aesthetics on (perceived) usability with some studies reporting some effect (Hartmann, Sutcliffe, & de Angeli, 2008; Lavie & Tractinsky, 2004); others found no

Table 10.1 An assortment of different authors' proposed dimensions for User Experience. Note the substantial difference in areas of concern: Bargas-Avila and Hornbæk (2011), Hassenzahl (2008a), and Bevan (2008) all have an obvious focus on the user's own perceived state and experience, whereas Robert and Lesage (2011b) consider perceived experience and context and psychophysical state, and Lim et al. (2008) take an almost fully instrumental view. Different, and somewhat orthogonal, dimensions can be obtained from various scientific traditions used to analyze user experience, such as the nine disciplines identified by Obrist et al. (2012b).

Robert and Lesage (2011b)	Cognitive, functional, physical, psychological, perceptive, social
Bargas-Avila and Hornbæk (2011)	Generic UX, affect and emotion, enjoyment and fun, aesthetics and appeal, hedonic quality, engagement and flow, motivation, enchantment, frustration, other constructs
Hassenzahl (2008a)	Pragmatic quality, hedonic quality—stimulation, hedonic quality—identity, attractiveness
Bevan (2008)	Likability, pleasure, comfort, trust
Lim et al. (2008)	Interface, visual, tactile, content, function, performance

significant effect (Hassenzahl, 2008a; van Schaik & Ling, 2009), and one reported the "inverse" effect that usability may have effect on postuse perceived aesthetics (Tuch, Roth, Hornbæk, Opwis, & Bargas-Avila, 2012).

User Experience Design and Process

Practical guidelines abound for *designing* artifacts that provide beneficial User Experience, or integrates a focus on User Experience in the design process.

While these are often of great use in designing products that have both high usability and good User Experience, much of the literature comprises a laundry list of best practices, design patterns, and design guidelines, often for visual design of interactable UI elements, rather than User Experience in the sense discussed above and in the scientific literature. These guidelines occasionally have magazinelike qualities and employ "User Experience" synonymously with either user experience, or with user-centered design in general.

Some recent university textbooks adopt User Experience as a term covering all aspects of design, and evaluation of interactive technology. For example, *The UX Book* (Hartson & Pyla, 2012) covers process, design, and evaluation with emphasis on practical tasks, yet references and grounds the tasks in the scientific literature. This approach of *user experience design* is a proper subfield of the more general—and substantially broader—notion of *user-centered design*.

Commercial enterprises may struggle to incorporate User Experience in design or production processes (Innes, 2007; Rosenberg, 2007), and a good number of case studies exist, for example Budwig, Jeong, and Kelkar (2009) have considered the challenges of combining user experience design with agile software development methods such as Scrum, as have Law and Lárusdóttir (2015, p. 584) with the strong conclusion that several agile methods are "not favorable for UX work in practice," and Leadley, Pao, and Douglas (2005) report on the significant psychological and organizational challenges in instilling a "User Experience culture" in a traditional development company.

Recent advances have tried to tackle such challenges by providing practitioners and enterprises with methods to properly incorporate User Experience in the design process; for example, Roto, Väänänen-Vainio-Mattila, Law, and Vermeeren (2009) chart the different needs of stakeholders in an enterprise and describe the need for distinct kinds of User Experience evaluation for different types of products, and Kaasinen et al. (2015) provide five approaches that in total cover the perspectives of different stakeholders in the enterprise, and a methodology to use them together.

User Experience Evaluation and Measurement

Distinct from *design* methodology and *design* processes for incorporating User Experience as part of product development is the empirical evaluation of User Experience aspects of products or prototypes. As elsewhere in human-computer interaction, evaluation as part of the design process provides feedback for designers usable for improving designs in later iterations, or for choosing between alternative candidate products. Similarly, for researchers, evaluation offers opportunities to define and validate evaluation criteria and provide manifest empirical evidence for the relations between design decisions and user's experience and satisfaction with the artifacts produced.

As with the plethora of extant definitions or descriptions of User Experience, there is no consolidated account of such empirical evaluation.

The difficulties in producing such an account are compounded by the fact that User Experience is multifaceted: it contains dimensions where the validity of empirical measurement, or at least the proxies used in place of dimensions, is still debated—for example affect and emotion, evident among both researchers and field practitioners.

Law, van Schaik, and Roto (2014, p. 526) conducted semistructured interviews with 10 User Experience researchers and one User Experience professional, finding "skepticism as well as ambivalence towards User Experience measures and shared anecdotes related to such measures in different contexts"; in a follow-up survey study among HCI researchers and professionals, the 170 answers showed a more nuanced attitude, and Law further delved into the "tension" between different schools of thought devoted to empirical measurement of User Experience, in particular the qualitative design-based approaches (which is most prevalent in existing User Experience research) and the quantitative model-based approach (Law, 2011).

Some attempts at producing general, workable evaluation frameworks for User Experience measurement exist (Forlizzi, 2007; Oppelaar, Hennipman, & van der Veer, 2008), but these have been almost exclusively confined to theory. In contrast, the toolbox of practical measures is—as may be expected by the lack of consensus among experts—sprawling. In a 2010 survey, Vermeeren et al. (2010) identified 96 methods for User Experience evaluation, and categorized them according to 10 nominal variables and eight further information categories (e.g., "name," "product development phase," and the type of applications or design that the methods can be used for).

Methods for evaluation can be roughly partitioned according to when they can be used in the development process (e.g., as evaluation of early prototypes), the amount of user interaction observed (e.g. momentary versus protracted use, single versus multiple users), and the type (e.g., qualitative versus quantitative), means (e.g., video observation, questionnaires, biometrics), and topos (lab versus field) of data collection. In addition, methods may be categorized according to their scientific provenance (e.g., sociology versus computer science) and the User Experience dimensions they concern. Again, there seems to be no consolidated overview, although the categorization of Vermeeren et al. (2010) is a good point of entry.

As is known from other fields, there is a clear tradeoff between observing users interact with artifacts in the field (resource-consuming) versus observing them in a decontextualized situation such as in a lab. Vermeeren et al. (2010) found that the number of such measures is—as may be expected by the lack of consensus among experts—sprawling. In their survey, Vermeeren et al. found an almost equal number of lab and field methods in use, but contend that field methods, due to their time-consuming nature, may be less used in industry, and that the majority of such methods are instead in use in academia.

Some attempts have been made at evaluating User Experience, or extracting information related to User Experience, without direct observation of users, for instance relying on user narratives concerning their experience (typically a very low-cost solution if narratives are already available), for example using machine learning techniques, but results have shown that while this some useful information can be obtained this way, it cannot be expected to supplant traditional methods (Hedegaard & Simonsen, 2013, 2014).

Some of the User Experience dimensions that have no counterpart in traditional usability research have been the focus of development of specialized methods; we mention a few: for *aesthetical* dimensions, most work is based on psychological measures, see Hartmann, Sutcliffe, and de Angeli (2008), Hassenzahl, Lindgaard, Platz, and Tractinsky (2008), and Lavie and Tractinsky (2004) for overviews and examples. For examples based on *affective* or *emotional* dimensions, see Hazlett and Benedek (2007); Isbister and Höök (2007), and Isbister, Höök, Laaksolahti, and Sharp (2007); for still further examples see work based on Lang's Self-Assessment-Manikin (SAM) scale (Lang, 1980), which remains in widespread use.

A number of concrete instruments in common use in studies measure dimensions of User Experience. For example, AttrakDiff (Hassenzahl, 2008a) allows practitioners to measure several hedonic aspects, attractiveness as well as usability aspects ("pragmatic quality"). In addition, standard instruments that measure more generic aspects of the User Experience may be used, for example as in Borsci, Federici, Bacci, Gnaldi, and Bartolucci (2015).

A challenge is that users' experiences with an artifact are dynamic and may change over time as users settle into using the artifact, or are exposed to alternative ones. There are several theoretical models aiming at understanding, and predicting, the development of User Experience over time. For example, Karapanos, Zimmermann, Forlizzi, and Martens (2009) chart how different types of experiences change over time, from orientation over incorporation to identification, and the ContinUE model of Pohlmeyer, Hecht, and Blessing (2009) posits five distinct phases from anticipation to retrospective and describes their relationships and characteristics. Similarly, Kujala, Roto, Väänänen-Vainio-Mattila, Karapanos, and Sinnelä (2011b) derive a model of User Experience over time that emphasizes more aspects of experience over time, including learning and engagement, and difference between momentary and summarized experience.

From a practical perspective, there are fewer studies and methods available; examples provided by Kujala et al. (2011a, 2011b) describe the problems in properly identifying hedonic aspects of dynamic, long-term use, and develop a cost-effective method to let users retrospectively report their experiences over time.

Finally, there is the issue of domain-dependence: while the general principles of User Experience, as described in the literature, is sufficiently general to be technology independent—indeed, it may be applied with equal validity to a Neolithic worker's kiln and to a modern astronaut's instrument board—there may be different design concerns affecting use of technological artifacts across different fields, and there may be specific types of use, or interaction possibilities, which are more important in some products than in others. Indeed, this has been confirmed by studies investigating User Experience in mobile cloud photo sharing (Vartiainen & Väänänen-Vainio-Mattila, 2010), mobile Internet use (Kaasinen, Roto, Roloff, Väänänen-Vainio-Mattila, & Vainio, 2009), Web services (Väänänen-Vainio-Mattila, & Wäljas, 2009), ubiquitous computing (Väänänen-Vainio-Mattila, Olson, & Häkkilä, 2015), and others.

What is a Practitioner to Do?

There is no standard flowchart for choosing an appropriate evaluation method for User Experience, and there is a plethora of methods in the literature that, in the words of Bargas-Avila and Hornbæk (2011, p. 2696) "overemphasize their methodological stance to the extent of damaging research quality … [and] do not report interview questions or protocols, rarely describe data analysis methods, focus mostly on generic UX." In addition, proponents of different approaches to *when* in the development process, and *what* to measure, and *how* to analyze the results, may take strong philosophical stances in describing their own view.

However, there are standard considerations to make when planning and scoping User Experience evaluations. Most of these considerations are common to all kinds of evaluation in human-computer interaction, but require special attention, or variation, when considering User Experience. A very brief account of these considerations follows below:

- What is the purpose of the evaluation? Explore design concepts? Measure specific dimensions of User Experience? Compare the User Experience of different artifacts or prototypes? A probe to see what aspects of the User Experience should undergo full evaluation?
- What is the scope of the artifact under consideration? Is part or the entire artifact evaluated? Are specific uses or workflows evaluated?
- What is the level of functionality tested? Nonfunctional prototype? Interactive prototype? Production code?

- Which dimensions of User Experience are to be evaluated? Aesthetics? Hedonics? Satisfaction? Something else?
- Should the evaluation be summative ("the hedonic quality of artifact A is 15% higher than that of artifact B") or formative ("the main problem with the hedonic quality of artifact A is the lack of a 'retry' functionality")?
- What are the resources available to carry out the evaluation?
- Does the ambient organization or community have demands on the type of evaluation? Is proven validity of the evaluation method a requirement?

Depending on the answers to the above questions, it may be that another notion of evaluation is needed (e.g., a traditional usability test), but it is likely that even specific answers will lead to further questions to be answered. Again, these questions occur elsewhere in human-computer interaction as well but may require special consideration for User Experience:

- What is the duration of the observation period? Is the artifact evaluated after protracted use? Several times? With what frequency? Are users' momentary experiences solicited, or their experiences over time?
- What is the topos of evaluations? Are evaluations done locally? In a lab? Remotely?
- What data are available for evaluation? User observations? User statements? Metadata? User Narratives? Physiological measurements of users?

The answers to these questions may sometimes be dictated by the types of evaluative instruments available; for example, instruments for measuring affective response may require extensive user presence in a lab (for instance, for measuring psychophysiological responses), or self-reported answers by users if questionnaires are used. A full mapping that associates answers to the above questions to an appropriate set of evaluation methods and instruments does not exist in the literature, and is beyond the scope of the present text. However, valuable resources do exist: summaries and categorizations of methods *in use* by both academics and industry experts are available (Bargas-Avila & Hornbæk, 2011; Obrist et al., 2009; Vermeeren et al., 2010).

For longitudinal studies with the aim of evaluating User Experience of long-term use, and the potential shifts and changes in User Experience, Kujala et al. (2011b) propose the "UX curve," a method for cost-effectively eliciting long-term User Experience information by properly guiding users through retrospective reporting.

References

Alben, L (1996). Quality of experience: Defining the criteria for effective interaction design. *Interactions, 3*(3), 11–15.

Bargas-Avila, J., & Hornbæk, K. (2011). Old wine in new bottles or novel challenges: A critical analysis of empirical studies of user experience. In D. Tan, S. Amershi, B. Begole, W. A. Kellogg & M. Tungare (Eds.), *Proceedings of the International Conference on Human Factors in Computing Systems, CHI 2011* (pp. 2689–2698). New York, NY: ACM.

Bevan, N. (2008). Classifying and selecting UX and usability measures. In *Proceedings of the International Workshop on Meaningful Measures: Valid User Experience Measurement* (pp. 13–18). Toulouse: Institute of Research in Informatics of Toulouse (IRIT).

Borsci, S., Federici, S., Bacci, S., Gnaldi, M., & Bartolucci, F. (2015). Assessing user satisfaction in the era of user experience: Comparison of the SUS, UMUX, and UMUX-LITE as a function of product experience. *International Journal of Human-Computer Interaction, 31*(8), 484–495.

Budwig, M., Jeong, S., & Kelkar, K. (2009). When user experience met agile: A case study. In D. R. Olsen Jr., R. B. Arthur, K. Hinckley, M. R. Morris, S. E. Hudson & S. Greenberg (Eds.), *Proceedings of the International Conference on Human Factors in Computing Systems, CHI 2009 (Extended Abstracts)* (pp. 3075–3084). New York, NY: ACM.

Czikszentmihaliy, M. (1997). *Flow and the Psychology of Discovery and Invention*. New York, NY: Harper Perennial.

Forlizzi, J. (2007). The product ecology: Understanding social product use and supporting design culture. *International Journal of Design, 2*(1), 11–20.

Forlizzi, J., & Battarbee, K. (2004). Understanding experience in interactive systems. In D. Benyon, P. Moody, D. Gruen & I. McAra-William (Eds.), *Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, DIS '04* (pp. 261–268). New York, NY: ACM.

Hartmann, J., Sutcliffe, A., & de Angeli, A. (2008). Towards a theory of user judgment of aesthetics and user interface quality. *ACM Transactions on Computer-Human Interaction, 15*(4), article no. 15.

Hartson, R., & Pyla, P. S. (2012). *The UX book*. Burlington, MA: Morgan Kaufmann.

Hassenzahl, M. (2004). The thing and I: Understanding the relationship between user and product. In M. A. Blythe, K. Overbeeke, A. F. Monk & P. C. Wright (Eds.), *Funology: From usability to enjoyment* (pp. 31–42). Dordrecht: Kluwer Academic Publishers.

Hassenzahl, M. (2008a). The interplay of beauty, goodness, and usability in interactive products. *Human-Computer Interaction, 19*(4), 319–349.

Hassenzahl, M. (2008b). User experience (UX): Towards an experiential perspective on product quality. In E. Brangier, G. Michel, J. M. C. Bastien & N. Carbonell, *Proceedings of the 20th International Conference of the Association Francophone d'Interaction Homme-Machine, IHS '08* (pp. 11–15). New York, NY: ACM.

Hassenzahl, M., Lindgaard, G., Platz, A. & Tractinsky (2008). The study of visual aesthetics in human-computer interaction. *Dagstuhl Seminar Proceedings* (Volume 08292). Schloss Dagstuhl: Leibniz-Zentrum für Informatik.

Hassenzahl, M., & Roto, V. (2007). Being and doing: A perspective on user experience and its measurement. *Interfaces, 72*, 10–12.

Hassenzahl, M., Schobel, M., & Trautmann, T. (2008). How motivational orientation influences the evaluation and choice of hedonic and pragmatic interactive products: The role of regulatory focus. *Interacting with computers, 20*(4–5), 473–479.

Hassenzahl, M., & Tractinsky, N. (2006). User experience: A research agenda. *Behaviour and Information Technology, 25*(2), 91–97.

Hazlett, R. L., & Benedek, J. (2007). Measuring emotional valence to understand the user's experience of software. *International Journal of Human-Computer Studies, 65*(4), 306–314.

Hedegaard, S., & Simonsen, J. G. (2013). Extracting usability and user experience information from online user reviews. In W. E. Mackay, S. A. Brewster & S. Bødker (Eds.), *Proceedings of the International Conference on Human Factors in Computing Systems, CHI '13* (pp. 2089–2098). New York, NY: ACM.

Hedegaard, S. & Simonsen, J. G. (2014). Mining until it hurts: Automatic extraction of usability issues from online reviews compared to traditional usability evaluation. In V. Roto, J. Häkkilä, K. Väänänen-Vainio-Mattila, O. Juhlin, T. Olsson & E. T. Hvannberg (Eds.), *Proceedings of the Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational, NordiCHI '13* (pp. 157–166). New York, NY: ACM.

Innes, J. (2007). Defining the user experience function: Innovation through organizational design. *Interactions, 14*(3), 36–37.

Isbister, K. & Höök, K. (2007). Evaluating affective interactions. *International Journal of Human-Computer Studies, 65*(4), 273–274.

Isbister, K., Höök, K., Laaksolahti, J. & Sharp, M. (2007). The sensual evaluation instrument: Developing a trans-cultural self-report measure of affect. *International Journal of Human-Computer Studies, 65*(4), 315–328.

Kaasinen, E., Roto, V., Hakulinen, J., Heimonen, T., Jokinen, J. P. P., Karvonen, H.,… Turunen, M. (2015). Defining user experience goals to guide the design of industrial systems. *Behaviour and IT, 34*(10), 976–991.

Kaasinen, E., Roto, V., Roloff, K., Väänänen-Vainio-Mattila, K. & Vainio, T. (2009). User experience of mobile Internet: Analysis and recommendations. *International Journal of Mobile HCI, 1*(4), 4–23.

Karapanos, E., Zimmerman, J., Forlizzi, J. & Martens, J.-B. (2009). User experience over time: An initial framework. In D. R. Olsen Jr., R. B. Arthur, K. Hinckley, M. R. Morris, S.E. Hudson & S. Greenberg (Eds.), *Proceedings of the International Conference on Human Factors in Computing Systems, CHI 2009* (pp. 729–738). New York, NY: ACM.

Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., & Sinnelä, A. (2011a). Identifying hedonic factors in long-term user experience. In A. Deserti, F. Zurlo & F. Rizzo (Eds.), *Proceedings of DPPI '11, Designing Pleasurable Products and Interfaces* (pp. 1–17). New York, NY: ACM.

Kujala, S., Roto, V., Väänänen-Vainio-Mattila, K., Karapanos, E., & Sinnelä, A. (2011b). UX curve: A method for evaluating long-term user experience. *Interacting with Computers, 23*(5), 473–483.

Lang, P. J. (1980). Behavioral treatment and bio-behavioral assessment: computer applications. In J. B. Sidowski, J. H. Johnson & T. A. Williams (Eds.), *Technology in Mental Health Care Delivery Systems* (pp. 119–137). New York, NY: Ablex.

Laurel, B. K. (1986). Interface as mimesis. In D. A. Norman & S. W. Draper (Eds.), *User centered systems design: New perspectives on human-computer interaction* (pp. 67–85). Hillsdale, NY: CRC.

Lavie, T., & Tractinsky, N. (2004). Assessing dimensions of perceived visual aesthetics of web sites. *International Journal of Human-Computer Studies, 60*(3), 269–298.

Law, E. L.-C. (2011). The measurability and predictability of user experience. In Paternò, F., Luyten, K. & Maurer, F. (Eds.), *Proceedings of the Symposium on Engineering Interactive Computing Systems, EICS '11* (pp. 1–10). New York, NY: ACM.

Law, E. L.-C., & Lárusdóttir, M. K. (2015). Whose experience do we care about? Analysis of the fitness of Scrum and Kanban to user experience. *International Journal of Human-Computer Interaction, 31*(9), 584–602.

Law, E. L.-C., & van Schaik, P. (2010). Editorial: Modeling user experience—an agenda for research and practice. *Interacting with Computers, 22*(5), 313–322.

Law, E. L.-C., van Schaik, P., & Roto, V. (2014). Attitudes towards user experience (UX) measurement. *International Journal of Human-Computer Studies, 72*(6), 526–541.

Leadley, B., Pao, H., & Douglas, S. (2005). Creating a user experience culture at a non-software company. In R. Anderson, B. Blau & J. Zapolski (Eds.), *Proceedings of the Conference on Designing for User Experience, DUX '05* (article no. 38). New York, NY: American Institute of Graphic Arts.

Lim, Y.-L., Donaldson, J., Jung, H., Kunz, B., Royer, D., Ramalingam, S.,…Stolterman, E. (2008). Emotional experience and interaction design. In C. Peter & R. Beale (Eds.), *Affect and emotion in human-computer interaction: From theory to applications* (pp. 116–129). Berlin: Springer-Verlag.

McCarthy, J., & Wright, P. (2007). *Technology as experience*. Cambridge, MA: MIT Press.

McCarthy, J., Wright, P., Wallace, J., & Dearden, A. (2006). The experience of enchantment in human-computer interaction. *Personal and Ubiquitous Computing, 10*(6), 369–378.

McNamara, N., & Kirakowski, J. (2006). Functionality, usability, and user experience: Three areas of concern. *Interactions, 13(6),* 26–28.

Ní Chongchúir, M., & McCarthy, J. (2008). The enchanting potential of technology: A dialogical case study of enchantment and the Internet. *Personal and Ubiquitous Computing, 12(*5*),* 401–409.

Norman, D. A., Miller, J., & Henderson, D. A. Jr. (1995). What you see, some of what's in the future, and how we go about doing it: HI at Apple Computer. In J. Miller, I. R. Katz, R. L. Mack & L. Marks (Eds.), *Proceedings of the International Conference on Human Factors in Computing Systems, CHI '95, Conference Companion: Mosaic of Creativity* (pp. 155). New York, NY: ACM.

Obrist, M., Roto, V., Law, E. L.-C., Väänänen-Vainio-Mattila, Vermeeren, A. P. O. S. & Buie, E. A. (2012a). Theories behind UX research and how they are used in practice. In J. A. Konstan, E. H. Chi & K. Höök (Eds.), *Proceedings of the International Conference on Human Factors in Computing Systems, CHI '12 (Extended abstracts)* (pp. 2751–2754). New York, NY: ACM.

Obrist, M., Roto, V., & Väänänen-Vainio-Mattila, K. (2009). User experience evaluation: Do you know which method to use? In D. R. Olsen Jr., R. B. Arthur, K. Hinckley, M. R. Morris, S. E. Hudson & S. Greenberg (Eds.), *Proceedings of the International Conference on Human Factors in Computing Systems, CHI 2009* (pp. 2763–2766). New York, NY: ACM.

Obrist, M., Roto, V., Vermeeren, A., Väänänen-Vainio-Mattila, K., Law, E. L.-C., & Kuutti, K. (2012b). In search of theoretical foundations for UX research and practice. In J. A. Konstan, E. H. Chi & K. Höök (Eds.), *Proceedings of the International Conference on Human Factors in Computing Systems, CHI '12 (Extended abstracts)* (pp. 1979–1984). New York, NY: ACM.

Oppelaar, E.-J. R. G., Hennipman, E.-J. & van der Veer, G. C. (2008). Experience design for dummies. In J. Abascal, I. Fajardo, & I. Oakley (Eds.), *Proceedings of the European Conference on Cognitive Ergonomics, ECCE '08* (article no. 27). New York, NY: ACM.

Overbeeke, K., Djajadiningrat, T., Hummels, C., & Wensveen, S. (2000). Beauty in usability: Forget about ease of use! In W. Green & P. Jordan (Eds.), *Pleasure with products, beyond usability* (pp. 9–18). London: Taylor & Francis.

Pohlmeyer, A. R., Hecht, M., & Blessing, L. (2009). Experience lifecycle model ContinUE [continuous user experience]. In C. Stössel & C. Clemens (Eds.), *Der Mensch im Mittepunkt technischer Systeme. Fortschritt-Berichter VDI Reihe 22, vol. 29* (pp. 314–317). Düsseldorf: VDI Verlag GmbH.

Robert, J.-M. & Lesage, A. (2011a). Designing and evaluating user experience. In G. A. Boy (Ed.), *The handbook of human-machine interaction* (pp. 321–338). Farnham, Ashgate.

Robert, J.-M. & Lesage, A. (2011b). From usability to user experience. In G. A. Boy (Ed.), *The handbook of human-machine interaction* (pp. 303–320). Farnham, Ashgate.

Rosenberg, D. (2007). Introducing the 360 deg. view of UX management. *Interactions, 14*(3), 22–24.

Ross, P. R., Overbeeke, C. J., Wensveen, S. A. & Hummels, C. M. (2008). A designerly critique on enchantment. *Personal and Ubiquitous Computing, 12*(5), 359–371.

Roto, V., Väänänen-Vainio-Mattila, K., Law, E. L.-C., & Vermeeren, A. (2009). User experience evaluation methods in product development (UXEM09). In T. Gross, J. Gulliksen, P. Kotz, L. Oestreicher, P. Palanque, R. Prates, & M. Winckler (Eds.), *Proceedings of the Conference on Human-Computer Interaction, INTERACT '09* (pp. 981–982). Berlin: Springer-Verlag.

Scherer, K. (2005). What are emotions? And how can they be measured? *Social Science Information, 44*(4), 695–729.

Tuch, A. N., Roth, S. P., Hornbæk, K., Opwis, K., & Bargas-Avila, J. A. (2012). Is beautiful really usable? Toward understanding the relation between usability, aesthetics, and affect in HCI. *Computers in Human Behavior, 28*(5), 1596–1607.

Van Schaik, P., & Ling, K. (2009). The role of context in perceptions of the aesthetics of web pages over time. *International Journal of Human-Computer Studies, 67*(1), 79–89.

Väänänen-Vainio-Mattila, K., Olson, T., & Häkkilä, J. (2015). Towards deeper understanding of user experience with ubiquitous computing systems: Systematic literature review and design framework. In J. Abascal, S. Barbosa, M. Fetter, T. Gross, P. Palanque & M. Winckler (Eds.), *Proceedings of the Conference on Human-Computer Interaction, INTERACT '15* (pp. 384–401). Berlin: Springer-Verlag.

Väänänen-Vainio-Mattila, K., & Wäljas, M. (2009). Development of evaluation heuristics for web service user experience. In D. R. Olsen Jr., R. B. Arthur, K. Hinckley, M. R. Morris, S. E. Hudson & S. Greenberg (Eds.), *Proceedings of the International Conference on Human Factors in Computing Systems, CHI 2009* (pp. 3679–3684). New York, NY: ACM.

Vartiainen, E., & Väänänen-Vainio-Mattila, K. (2010). User experience of mobile photo sharing in the cloud. In M. C. Angelides, L. Lambrinos, M. Rohs & E. Rukzio (Eds.), *Proceedings of the International Conference on Mobile and Ubiquitous Multimedia, MUM '10* (p. 4). New York, NY: ACM.

Vermeeren, A. P. O. S., Law, E. L.-C., Roto, V., Obrist, M., Hoonhout, J., & Väänänen-Vainio-Mattila, K. (2010). User experience evaluation methods: Current state and development needs. In A. Blandford & J. Gulliksen (Eds.), *Proceedings of the Nordic Conference on Human-Computer Interaction, NordiCHI '10* (pp. 521–530). New York, NY: ACM.

Whiteside, J., & Wixon, D. (1987). The dialectic of usability engineering. In G. Salvendy, S. L. Sauter, J. J. Hurrel Jr. (Eds.), *Proceedings of the Conference on Human-Computer Interaction, INTERACT '87* (pp. 17–20). Amsterdam: Elsevier.

Wiklund-Engblom, A., Hassenzahl, M., Bengs, A., & Sperring, S. (2009). What needs tell us about user experience. In *Proceedings of the Conference on Human-Computer Interaction, INTERACT '09* (pp. 666–668). Berlin: Springer-Verlag.

Wright, P., McCarthy, J., & Meekison, L. (2004). Making sense of experience. In M. A. Blythe, K. Overbeeke, A. F. Monk & P. C. Wright (Eds.), *Funology: From usability to enjoyment* (pp. 43–53). Dordrecht, Kluwer Academic Publishers.

11

Task Load and Stress

Julien Epps

Introduction

A "task" is a fundamental unit of human activity from the perspective of human-computer interaction. While the definition of a human task is not straightforward, the notion of task load is somewhat easier to define. If we observe someone completing any kind of task, we can decide fairly easily, for example, how engaged or loaded they are, whether to interrupt them, what kind of assistance they might need, and when they are ready for the next task. Computing systems can easily make these decisions about task progress and load for *computing* tasks in real time (e.g. Task Manager in Windows); however, they are currently incapable of making these decisions about *human* tasks in general.

Arguably the most commonly investigated form of task load is mental load, which is an artefact of the limited processing capacity of the human brain. Working memory is an important concept for expressing this limited capacity, and is defined as a system that provides temporary storage and maintenance of stored information, and that provides an interface with perception, long-term memory, and action (Baddeley, 2003). Resource theory shares the notion of a finite pool of attentional resources, which can be engaged in one or more tasks, up to a capacity limit (Kahneman, 1973). Related to this is the more general construct of cognitive load, which is broadly defined as the load imposed by a particular task on the cognitive system (Paas et al., 2003). Use of the term "mental load" in the literature often refers to or includes either working memory load or cognitive load; however, task load can often also include physical demand, time pressure, performance, effort, fatigue, frustration, or perceptual demand, for example.

The notion of physical load has received less attention in HCI, presumably because of the limited range of applications to date; however as computing becomes more mobile and assistive, physical load may become increasingly important. Despite this, the notions of physical load and its measurement have also been investigated for some time (e.g. Borg, 1990). Physical and mental load can both be described in terms of the level of task demand and the impact on the user (Parasuraman and Hancock, 2001; Young, Brookhuis, Wickens, & Hancock, 2015).

The Wiley Handbook of Human Computer Interaction, Volume 1, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

Stress is most commonly recognized as a response that degrades performance capability (Hancock & Warm, 1989) or disturbs normal function (Tepas & Price, 2001). McGrath (1976) explained the stress response as an interaction between perceived demand, perceived coping ability, and perceived importance of coping. In HCI, very often notions of task load and stress are explicitly or implicitly linked with notions of task performance, and in many applications, the latter is a key motivation for investigating the former. Hancock and Warm (1989) describe stress in terms of states of relative stability, transition towards degradation, and degradation towards collapse, i.e. all in terms of task performance. Depending on the type of stress, there may be some similarities between stress and task load. For example, stress has been found in many studies to impair working memory function.

By way of distinguishing task load and stress, Gaillard (1993) explains high cognitive load as a normal, healthy state in which additional mental effort and/or modified strategies are engaged to adapt to the demands of the task, while explaining stress as a disturbed equilibrium characterized by inefficient behavior, overreactivity, and possibly adverse physiological effects. Hollien (1980) refers to a stress as the response to some sort of threat. In a stressed state, additional mental effort and/or modified strategies may be ineffective in dealing with the task or situation at hand (Gaillard, 1993). Coping strategies also differ: during cognitive load, strategies have been experimentally observed to be more focused on the task, while during stress, strategies may be more emotion focused.

Importantly, however, high workload can contribute to the development of stress symptoms, particularly psychological stress. Tepas and Price (2001) associate stress with concepts like anxiety, burnout, coping, exhaustion, exposure, fatigue, hardiness, repetitiveness, strain, stressor, and tension, as well as mental load. Some assessment scales include both mental load and psychological stress—for example, the Subjective Workload Assessment Technique (Reid & Nygren, 1988).

Considering these various concepts and definitions, clearly there is no universal definition of task load. It is also clear that task load is a multidimensional concept, and that to understand it in a more comprehensive manner, it may be necessary to seek inspiration from taxonomies of human tasks. Indeed, other forms of task load than mental and physical load may take on increased importance with the changing role of computing. Stress is similarly not straightforward to define, but may be distinguishable from high task load in terms of its adverse physiological and psychological response.

Quantifying Task Load and Stress

From theoretical to functional descriptions

In order that the notions of task load and stress can be applicable to HCI, it is necessary to describe them and their intensity. Borg (1990) explains the problems associated with a lack of fundamental units for measuring perceptual intensities. This lack of units means that even when numerical scales have been proposed for task load or stress intensity, the numbers have an ordinal rather than a numerical interpretation, e.g. a scale rating of "4" indicates a greater intensity than a rating of "2," but is by no means necessarily double the intensity of "2." In some applications, researchers have adopted a more categorical approach, for example "low," "medium," and "high" load (e.g. Chen et al., 2012). Questions remain as to whether task load and stress should be measured on a ration, interval, or ordinal scale. The interval scale is sufficient for use with parametric statistics, and the ordinal for nonparametric statistics. Two requirements of descriptions of task load and stress should be that they permit interindividual comparisons, and that they are "meaningful" (Borg, 1990). Many approaches to describing task load and stress have not attempted or achieved the former, and it is likely that scales or categorical representations will be primarily meaningful within individuals, due to fundamental differences between individuals (Jarrold & Towse, 2006). Task load or stressor intensity are taken to refer to a particular period of time, and although some of the literature is not precise about this aspect, broadly speaking the unit of time is usually the duration of the task being completed, often under the assumption that there is a single task, or at least a single primary task.

Two example assessment scales

The Subjective Workload Assessment Technique (SWAT) (Reid & Nygren, 1988) comprises three scales: time load, mental effort load, and psychological stress load. Each scale has three ordinal levels, with descriptors, for example "Very little conscious mental effort or concentration required" through to "Extensive mental effort and concentration are necessary," which help to anchor the ratings applied by users during assessment.

The NASA Task Load Index (TLX) as commonly used comprises six scales: mental demand, physical demand, temporal demand, performance, effort, and frustration (Hart & Staveland, 1988). These are typically represented with up to 21 gradations, with descriptors of "very low" and "very high" at the ends of the scales. Originally, a further four scales were proposed, including stress, fatigue, activity type, and task difficulty.

The SWAT and TLX, both commonly used scales, are multidimensional, and through their various dimensions reflect at least some of the diversity of different types of load that may occur during a task. The multiple dimensions also offer some diagnosticity, although other approaches such as the workload profile may be more effective in this respect (Rubio, Diaz, Martin, & Puente, 2004). Given that many task-load scales are multidimensional, one question that arises is how to combine them if only a single measure of task load is required. This has been examined by Hendy, Hamilton, & Landry (1993), who found that it is sufficient to combine individual scales in an unweighted manner.

Task load, or task complexity, as quantified by these scales can have different possible origins. Ruiz (2011) categorizes these as content complexity, representational complexity, and user or learner characteristics. Content complexity or intrinsic load (Paas et al., 2003) refers to the intrinsic demand of the task itself, and is related to the number of atomic contributing subtasks that must be completed and temporary memory required to satisfy the overall task. Representational complexity or extraneous load (Paas et al., 2003) refers to the format in which the task is presented and the environment in which the task must be completed, and can be interpreted as being related to how "well designed" a computing interface is. User or learner characteristics, or germane load (Paas et al., 2003), account for the fact that a task perceived to be difficult for a novice may be simple for an expert.

One of the SWAT scales, psychological stress, describes low stress as when "little confusion, risk, frustration, anxiety exists and can be easily accommodated," and this is contrasted with "High to very intense stress due to confusion, frustration or anxiety. High to extreme self-determination and self-control required" (Reid & Nygren, 1988). The perceived stress scale (Cohen, Kamarck, & Mermelstein, 1983) is an alternative, subjective questionnaire very widely used outside of HCI, however its questions are long term, referring to "over the last month," and a short-term equivalent suitable for HCI tasks may yet be needed.

Investigating Task Load and Stress

Induction methods

In order to study task load, many researchers have constructed experiments to induce different levels of load in a controlled manner. Some experiments have been conducted in context—in the actual application environment or in a realistic but simulated version of this—while others have been conducted using elementary cognitive tasks, often in a laboratory setting. An advantage of the former is greater realism, while advantages of the latter include significantly more precise control over the task type and easier comparison with related literature.

Among different possible induction methods, there is a choice between dual-task methods and primary task methods. Dual-task methods, used for example by O'Donnell and Eggemeier (1986), require participants to perform a secondary task in parallel with the primary task (i.e. the one under study), and performance in the secondary task is used as an indicator of task load, as a measure of remaining resources not utilized by the primary task. Imposing a secondary task on participants may not be feasible in many practical experimental settings, and anecdotal evidence from some studies suggests that it can be difficult to keep participants engaged in both tasks (the secondary task can become neglected) (Wilson & Eggemeier, 1991), perhaps explaining why many studies use a primary task only.

If task-load intensity is to be varied over multiple levels, then the induction method should be capable of inducing a potentially large number of distinct load levels, with a minimal number of parameter changes between adjacent induced levels. For example, arithmetic tasks satisfy this criterion reasonably well, and have been used to induce up to seven distinct load levels (e.g. Zarjam, Epps, Chen, & Lovell, 2013); however between adjacent load levels two parameters may need to be adjusted: both the number of digits to be added (for example) and the number of carries. Another design choice for induction methods is between visual and verbal task presentation (Klingner, Tversky, & Hanrahan, 2011).

In terms of methods available for task load induction, many researchers have chosen memory span tasks. These present participants with a series of target stimuli (e.g. words or numbers) interleaved with a series of demanding secondary tasks (e.g. sentence comprehension), with the aim of measuring how well participants can recall the target stimuli from working memory at the conclusion (see e.g. Conway et al., 2005). Methods include reading span, word span, nonword span, letter span, speaking span, digit span, backward digit span, listening span, and counting span. Another working-memory induction method is the n-back task, which has been used extensively in cognitive neuroscience (see e.g. Owen, McMillan, Laird, & Bullmore, 2005). In this task, participants must respond whenever a target stimulus is presented that is the same as the one presented *n* trials earlier, with *n* typically selected as 1, 2, or 3.

Other types of task-load induction methods may be less focused on memory; for example the Stroop task (used e.g. by Yin et al., 2008), driving tasks (e.g. Strayer & William, 2001), elementary cognitive tasks focusing on visual perception or cognitive speed (Haapalainen, Kim, Forlizzi, & Dey, 2010), or the more perceptually demanding search task used by Chen, Epps, & Chen (2013). Among stress-induction methods, social evaluative threats have been used (e.g. Setz, Arnich, & Tröster, 2010), or psychosocial stress protocols like the Trier Social Stress Test, in which participants are asked to give a presentation in front of an audience (e.g. a job interview) and are asked a challenging and long mental arithmetic question. Online communities supporting a wide variety of standard and custom induction methods have recently begun to emerge, such as the Psychology Experiment Building Language (PEBL).

Practical and experimental considerations

A number of settings can have a very significant effect on experimental observations. When designing tasks, it is important to consider the possibility of overloading participants, as this may produce unexpected results (see below). It may be helpful to have an incentive for participants to remain engaged through any high load conditions. Under high load, participants will also naturally adopt strategies to reduce the load, including, for example, writing or counting on fingers, and instructions may be needed to prevent this. It is very desirable to employ some form of secondary validation of induced load levels, to ensure that they are distinct—this has been achieved using subjective rating methods, response time, and / or task performance (e.g. Chen & Epps, 2013; de Greef, Lafeber, Van Oostendorp, & Lindenberg, 2009). If time pressure is introduced, in some cases this may induce participant responses or behaviors that are not purely related to task load. As with other experiments of this kind, it is often advisable to use counterbalancing or other forms of task sequence randomization. Currently, there is no consensus on the relationship between task load, stress, and arousal, however some initial studies in this area (e.g. Chen & Epps, 2013) suggest emotional arousal as a potential confounder for task load studies. During protocols involving physiological measurements, it is common to allow several minutes for participants to adjust to their environment before beginning tasks, to establish a baseline.

The importance of pilot studies in this kind of empirical research cannot be overstated. Pilots offer the opportunity to check that load-induction methods are feasible, to verify that they induce the intended types and intensities of load, to observe what strategies participants adopt, and to elicit participant feedback on the tasks.

As the study of task load and stress matures, the use of both constrained, laboratory-style tasks and more realistic tasks within the same protocol and participant cohort can be expected. Although this will likely involve participants completing different types of tasks, which raises questions about whether any single measurement approach could be consistently applied across the tasks, it is needed to answer questions of how well the findings from the literature on laboratory-style tasks will generalize to more realistic task conditions, and whether laboratory-style tasks can be used to calibrate task load measurement systems for those conditions. Since task load and stress are only probabilistically linked to measures of these effects, it is desirable to record as many measures as possible (Larsen & Fredrickson, 1999), and indeed this is a trend in multimodal HCI.

Measures of Task Load and Stress

Arguably the concepts of task load and stress have little relevance to practical HCI systems unless they can be measured. However, in order to be useful, task-load measures should ideally be sensitive to task demand, specific to the type of task demand dimension being investigated, help to isolate the origins of the task demand, and not obstruct the performance of the user's task, and maintain these attributes consistently over time (O'Donnell & Eggemeier, 1986) and across different measurement contexts (Kramer, 1991). In surveying the current status and limitations of existing methods for task load and stress measurement, for many it may be more correct to say that they are estimating rather than measuring task load. Among the different categories discussed below, it is interesting to note that only one (analytical) is directly related to the intrinsic level of task demand; all other measures relate to the effect of the task on the user (Young et al., 2015). Also, only three (physiological, behavioral, and to some extent performance) can be completely automated.

Analytical

For a limited range of very specific kinds of tasks, it may be possible to determine a priori what the intrinsic load of a task is. Methods have been proposed for decomposing and characterizing arithmetic tasks by analyzing them for example in terms of element interactivity (Sweller, 1994), similarly to the manner in which the complexity of computing algorithms may be analyzed.

Subjective

Subjective measures comprise ratings on scales such as those mentioned above, usually provided by users immediately on finishing a task. These have been observed to provide very high correlations between repeated tasks at the same load level (Gopher & Donchin, 1986), and scales with up to nine gradations have been shown to provide statistical significance in distinguishing adjacent levels of load (Paas et al., 1994). Consequently, subjective measures have been very widely used and are essentially a de facto standard task-load measure. Although there are many scales, perhaps the most common is a single-dimension scale similar to the mental demand TLX scale. In Nygren's (1991) investigation of SWAT and TLX, he suggests that TLX be preferred as a general prediction approach, while SWAT is useful as a simplified psychological model of task-load judgment.

Subjective measures are a direct measure of the perception of the task load by the user, and are usually collected after the task has completed, which presents three limitations: (a) the ratings may be limited by individual bias; (b) users may not remember the load of the entire task well once it has completed (Young et al., 2015), and (c) there is no possibility for assessing the instantaneous workload during the task. Ratings from subjective measures may also be difficult to interpret or compare across different types of tasks (Gopher & Donchin, 1986).

Nygren (1991) raised concerns about the lack of crossvalidation studies and multi-trait, multimethod studies of subjective measurements, and the prevalence of studies that were not rigorously validated or provided few insights into the information processing system of individual users. Many of these concerns probably still stand today, and may be equally applicable to studies of physiological and behavioral measures.

Performance

Task load is often investigated in the context of task performance, or with this in mind, dating back to experimental work showing a relationship between arousal and task performance, known as the Yerkes–Dodson law or the "inverted U-curve." Although this law is still the subject of some debate, and notions of underload and overload are challenging to define precisely, the general characteristic of lower performance during either very low or very high task demands and higher performance during moderate task demands persists through to very recent reviews of the topic (Young et al., 2015). The characteristic is a fundamentally important motivation for human-computer interaction research in that it identifies an optimal region for human performance.

De Waard (1996) refers to low task demand as requiring *state-based effort* to remain engaged, and if the task is extremely monotonous this may require high levels of effort. This is represented by the leftmost part of Figure 11.1, and example tasks associated with this region may be related to vigilance and/or monitoring (e.g. long-distance driving or continuous operation of machinery). On the other hand, the rightmost parts of Figure 11.1 are associated with *task-based effort*, a

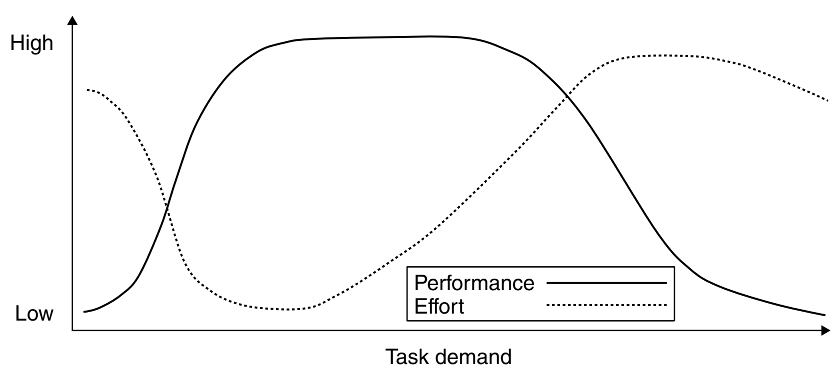


Figure 11.1 Conceptual model of performance as a function of the mental task demand (after De Waard, 1996). To the left, low task demand can be associated with low performance due to low user engagement with the task (underload), while on the right the decrease in performance with increasing task demand reflects increased demand on the user's constrained mental processing resources. Although not straightforward to define, overload is often associated with the higher task demand phase of the curve in which performance drops far from optimum, and in which task disengagement may occur.

characteristic that has been observed in many empirical studies (see for example Figure 6 in Zarjam, Epps, & Lovell, 2013).

Sources of dissociation between subjective load measures and performance are discussed in detail by Yeh and Wickens (1988). From a measurement perspective, the characteristic illustrated in Figure 11.1 demonstrates two important effects: (a) that a particular level of task performance may exist under relatively low and relatively high load conditions, i.e. performance cannot unambiguously indicate the level of spare capacity; and (b) when high load levels are induced, overload can occur.

Performance cannot be generalized across different task types, and in general computing systems, where users may be interacting with multiple computing applications and switching between them, it may be difficult or impossible to measure performance, even if individual computing applications each have associated performance measures. As a measure, performance seems best suited to very specific types of tasks, which are persistently the user's continuous primary source of attention and that allow performance measurements frequently relative to the task duration.

Physiological

Physiological measures of task load can be unobtrusive, do not require overt responses from the user, are often multidimensional, are increasingly inexpensive, and offer a degree of objectivity (Kramer, 1991; Wilson and Eggemeier, 1991). Physiological measures are also gaining attention in other parts of HCI, such as tracking physical health state. However, they are often sensitive to nontask variability and noise, need to be persistently physically correctly fitted, and must be carefully individually calibrated in order to be useful. Further, the fundamental dependence of mental workload on physiological changes is yet to be completely understood.

Perhaps the most convincing physiological signal in the task-load context is electroencephalography (EEG), which makes a direct observation of cortical activity, and can be sampled very fast, to potentially allow observation of rapid load changes (see e.g. Gevins et al., 1998). Example EEG features sensitive to task load are shown in Table 11.1; all are indicative of the expected higher cortical activity with increased task load, often in the frontal lobe area associated with higher cognitive function. Recent initial results suggest that EEG may have greater sensitivity to task load than subjective rating (Zarjam et al., 2015). By the nature of cortical activity and the spatial distribution of electrodes, EEG is task dependent, which is both a limitation (if a single task load measure is needed across different task types) and an opportunity (to localize the task type). EEG requires time and expertise to operate, and is suited to the laboratory context, although wireless EEG headsets may begin to change this. Evoked potential (ERP) has also been studied in conjunction with task load; however, this also requires a controlled and synchronized stimulus.

A number of physiological signals may be recorded that reflect autonomic nervous system (ANS) function, including pupil diameter, blink, heart rate, respiration, electrodermal activity, and hormone level, as shown in Table 11.1—these can be measured reasonably easily, although good sensor contact with the skin or body is needed. As an indicator of mental activity and arousal, galvanic skin response (GSR) or electrodermal activity (EDA) has been employed in a variety of different task

Table 11.1 Selected reported physiological and behavioral effects of increased mental task load.

Measure		*Effect*	*Example reference(s)*
EEG	Frontal midline theta rhythm	↑	Gevins et al. (1998)
	Parietocentral alpha rhythm	↓	Gevins et al. (1998)
	Low frequency band wavelet coefficient energy (frontal)	↑	Zarjam et al. (2015)
ERP	P300	↓	Kramer. (1991)
GSR	Mean	↑	Shi, Ruiz, Taib, Choi, and Chen (2007)
	Quantile values of peaks/min	↑	Setz et al. (2010)
Heart rate	Mean	↑	de Waard (1996) Wilson and Eggemeier (1991)
	Variability (HRV)/Median absolute deviation	↓	de Waard (1996) Haapalainen et al. (2010)
Heat flux	Raw value	↑	Haapalainen et al. (2010)
Respiration	Respiration rate	↑	Wilson and Eggemeier (1991)
Eye activity	Pupil dilation (increase from baseline)	↑	Peavler (1974) Chen and Epps (2013)
	Blink number, duration	↓	Peavler (1974) Chen et al. (2013)
	Fixation duration	↑	de Greef et al. (2009)
Speech	Duration-related spectral features	↑	Yin et al. (2008)
	Creaky voice quality	↑	Yap, Epps, Ambikairajah, and Choi (2015)
Mouse	Click pressure change rate	↓	Ikehara and Crosby (2005)
Writing	Minimum velocity	↓	Yu, Epps, and Chen (2011)
Accelerometry	Head/torso movement	↓	Makepeace and Epps (2015)

contexts, and increased skin conductance has been shown in many studies to be associated with higher mental load / or stress; however, it is also associated with other effects on the sympathetic nervous system and environmental factors. Respiration belts have often been used in recent studies to observe increases in respiration rate with mental load. Heart rate has the advantage that frequency components associated with distinct physiological control mechanisms can be isolated, but is affected by physical activity and other factors in addition to mental load. The rate of heat transfer has also shown promise for indicating high task load in one study (Haapalainen et al., 2010).

Pupil diameter has attracted significant attention in psychological research, and since its connection with memory load was established (Kahneman & Beatty, 1966) it has held ongoing interest as a measure of task load. Pupil diameter can be measured in laboratory settings using either desk-mounted or head-mounted infrared eye tracking equipment, and smaller, low-cost versions of the latter have also been employed recently (Chen et al., 2013). A complication of its use is the fact that the pupillary light reflex is much stronger than the effect due to changes in task load. For an overview of methods for extracting information from eye video, please refer to Chapter 21.

Behavioral

Behavioral measures of task load are more speculative than physiological, as they are measured further from the source of the mental effect of the task, but are among the least intrusive, most cost effective and most practical for integration into a human-computer interface. Sharing many of the same advantages and limitations as physiological measures, they are challenging to work with because the signal features usually contain effects unrelated to task load, and often these effects appear stronger than the task load effect. Chen et al. (2012) provide a more extensive introduction to behavioral methods as task-load indicators than the brief summary below.

Eye activity features other than pupil dilation, such as blink, fixation, and saccade, may be more physiological or behavioral, depending on the degree of voluntary control involved. Eye blink can be very robustly detected, and Siegle, Ichikawa, and Steinhauer (2008) showed that blink occurs mainly before and after tasks, supported by recent work (Chen et al., 2013).

Speech is a feasible measure whenever a microphone can be placed near the user, and although environmental noise and other factors remain a challenge, there are signs that two to three task load levels can be reliably distinguished. Interestingly, classification of cognitive load from speech has recently been the subject of an international datacentric evaluation effort—the ComParE Challenge (Schuller et al., 2014). Speech under stress has also received significant research attention (e.g. Hansen & Patil, 2007).

Sensors built into computing devices are constantly providing new opportunities for behavioral measures, often on a ubiquitous scale. Measures derived from pen sensors (Ruiz et al., 2010; Yu et al., 2011) and mouse pressure sensors (Ikehara & Crosby, 2005) have attracted research interest. An early head-based accelerometry study shows promise for indicating task transition, mental task load (Makepeace & Epps, 2015) and physical task load, although activity recognition (e.g. Bao & Intille, 2004) is more common in accelerometry research.

Future challenges

Psychometric properties such as reliability, validity, and sensitivity have not been established for many measures, particularly physiological and behavioral. This includes assessing test-retest reliability (e.g. Gevins et al., 1998), which is common in medicine and parts of psychology, but has not yet been studied for all measures across this field, and particularly not for newer measurement approaches.

It has been noted that different forms of subjective rating scales may be more applicable to different task types. As researchers work more with physiological and behavioral approaches, which can be applied in an automatic “always-on” mode, it seems likely that they will be applied across multiple different task types. It will be important to conduct large-scale studies which apply multiple different measurement approaches across a number of different task types, to understand how well the approaches generalize across task type, and to understand what the limitations of applying them to such diverse contexts are. This has begun; for example, Reinerman-Jones, Matthews, Barber, and Abich (2014) found that across four different task types and six different measurement approaches, although several approaches were sensitive to task load, no common latent factor could be identified.

Implications for Interface and System Design

Interface evaluation

Among many criteria for effective computer interfaces is the requirement to minimize the user's memory load (Oviatt, 2006), i.e. to reduce the amount of effort that must be exerted by a user in order to achieve a particular task, for example by reducing the representational complexity (see above) of the interface. Some researchers have implemented this as an evaluation approach using an EEG-based load measurement paradigm in conjunction with a think-aloud approach (Pike et al., 2012) or eye movements, in particular fixation (Lin, Zhang, & Watson, 2003).

Interface and system adaptation

Several researchers have pointed to the possibility of adapting interfaces or human-computer systems based on the user's instantaneous load level (e.g. Byrne and Parasuraman, 1996; de Greef et al., 2009), so that user task load is regulated as a function by shifting load to the computing system, or by simplifying the interface or task. Load-level awareness also gives rise to the notion of cognitive assistance (Siewiorek, Smailagic, & Starner, 2012), in which wearable computers try to assist the user in a timely manner with tools for offloading, after detecting high load.

Automatic estimation of task load and stress

Despite the number of papers written on task-load measures that can be automated, the number of fully functioning task load and stress estimation systems to date is small, and these are mainly research prototypes. Many authors have referred to the possibility of continuous measurement; however, to date virtually none have shown measurement that is truly continuous (usually reported on a per-task basis, where a "task" is often many seconds or minutes) and valid, perhaps due to the difficulty in determining a convincing ground truth on a fine temporal scale. Providing a meaningful measure of ground truth for practical load estimation systems also continues to be a challenge; however, it is likely that initial "calibration" tasks could be employed to help benchmark the level of load or stress for a given user when they commence using the system, or perhaps on a regular basis.

Automatic approaches to date have focused mainly on mental workload and to a lesser extent stress. The automatic assessment of other aspects of task load, including for example scales from well-established subjective workload assessment techniques, have received considerably less attention. The classification of tasks, however, has received significant attention: both physical—for example, accelerometer based (Bao & Intille, 2004) and mental (e.g. EEG-based). However, in all cases, this research is based on a limited, fixed set of predefined tasks, which will in general have limited applicability except in very constrained application environments. From this perspective, it seems that a more generic framework will be needed for task analysis, perhaps one that takes a taxonomy of human tasks into account.

The sheer ubiquity of diverse sensors in current and future computing devices strongly suggests that the above issues will be resolved in time.

Applications of task load and stress measurement

Critical environments: As suggested by Figure 11.1, many studies have found that mental task load is related to performance, and this has historically been a key motivation for studying task load in users operating in attention-critical environments. Safety-critical applications such as air-traffic control, emergency response, command and control, driving and operation of potentially dangerous machinery, have historically been a key area of task load research, and a recent survey of these by Young et al. (2015) gives a helpful overview.

Human-automation interaction: As computing devices become more mobile and more assistive in nature, many more joint human-machine systems will appear in daily life—an early example is GPS-enabled devices, which are widely used to assist driver navigation. In their seminal paper on human-automation interaction, Parasuraman, Sheridan, and Wickens (2000) decompose human information processing into four stages: sensory processing, perception / working memory, decision making, and response. Their model provides a structure for deciding which parts of a joint human-machine system should be automated; however, the model arguably relies on the machine being able to understand moment to moment what the user is capable of. From this perspective, an estimate of the user's task load is an important input to human-machine systems.

Automatic offloading assistance: In terms of task load and stress, computing devices already provide an important means to offload items that might otherwise occupy working memory (Siewiorek et al., 2012), for example it is now easy for users to set themselves reminders rather than remember key deadlines or events. However, it can be expected that this kind of support will become much richer and more instantaneous in future, for example glasses that automatically record photos or video of a situation if their sensors detect that a user's load is increasing rapidly. This is in line with Oviatt's (2006) vision of providing interfaces that help people think.

Task switching: Switching between tasks often occurs due to an interruption of some kind, and with the proliferation of email, landline / mobile / Internet telephony, instant messaging, text messaging and social media, interruptions are now a major part of the life of any professional. It is easy to guess the impact of such interruptions, and studies show that when primary task execution is interrupted, users require up to 27% more time to complete the tasks, commit twice the number of errors, experience up to 106% more annoyance, and experience twice the increase in anxiety than when interruptions are presented at the boundary between primary tasks (Bailey & Konstan, 2006). Humans know this, and are able to interrupt at opportune times, but computing systems do not. The notion of overload, be it information overload or work-related stress, also has important implications for health and wellbeing (Gaillard & Wientjes, 1994), and yet without an automatic method for task analysis (Chen & Vertegaal, 2004), it is difficult to even begin designing systems that could help to alleviate these problems and allow users to modify their work habits. It can also be observed that, in the data-analytic age, data on human task switching and task load is conspicuously absent, and approaches to automatic continuous task analysis that account for switching and load level are still emerging (Chen et al., 2013; Shen et al., 2009).

Measures of learning: As anticipated by the discussion of learner characteristics as a component of task complexity above, task load has interesting applications in the education domain, and manual (e.g. Paas et al., 1994) and objective (e.g. Brünken, Plass, & Leutner, 2003) computer-based measures of load during learning have been proposed for some time now. Martin (2014) notes that since task load measurements are relative, transient and influenced by other factors than the task, they may be better employed in a longitudinal context. This kind of longitudinal context was investigated by Ruiz et al. (2010) for cognitive training of elite athletes, and highlights a promising application area for task load: in cases where tasks can be sufficiently constrained, a reduction in load experienced can be expected as users learn the task and develop strategies to improve their task effectiveness, and this reduction should be observable with a sufficiently precise load measurement method, under the strict condition that the task type is not varied. Task load plays a role in the design of educational interfaces, and this is discussed in more detail by Hollender et al. (2010).

Conclusion and Future Directions

One of the great challenges facing the general area of task load and stress, in an era where machine learning is having a great impact, is how to automate the measurement of task load and stress in a manner that is accurate and meaningful for computing applications. Another challenge that has received little attention to date is to automate the detection of task transitions (historically accomplished manually—see, for example, Czerwinski, Horvitz, & Wilhite, 2004) in a general, mobile computing context, acknowledging that task load is more meaningful during tasks than during transitions, and that computing applications need to understand when humans change their primary task, at the very least to know when to interrupt them (Adamczyk & Bailey, 2004). If automation can be achieved, then (a) a much deeper partnership between humans and computing systems will be possible, taking much richer account of context; and (b) task load and stress can enter the age of big data, and a great number of new insights into human behavior may be gleaned. This may also converge well with newer computing applications designed to help humans offload their working memory and other types of task load, which can be expected to flourish as mobile computing brings information processing tasks closer to the user's attention for increasingly large proportions of every day.

Acknowledgments

This work was partly funded by the U.S. Army International Technology Center (Pacific).

References

Adamczyk, P. D., & Bailey, B. P. (2004). If not now, when? The effects of interruption at different moments within task execution. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 271–278). New York, NY: ACM.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews. Neuroscience, 4*(10), 829.

Bailey, B., & Konstan, J. (2006). On the need for attention-aware systems: Measuring effects of interruption on task performance, error rate, and affective state. *Computers in Human Behavior, 22*(4), 685–708.

Bao, L., & Intille, S. (2004). Activity recognition from user-annotated acceleration data. In *Proceedings of the International Conference on Pervasive Computing* (pp. 1–17). Heidelberg: Springer Lecture Notes in Computer Science.

Borg, G. (1990). Psychophysical scaling with applications in physical work and the perception of exertion. *Scandinavian Journal of Work, Environment and Health*, 55–58.

Brünken, R., Plass, J. L., & Leutner, D. (2003). Direct measurement of cognitive load in multimedia learning. *Educational Psychologist, 38*(1), 53–61.

Byrne, E. A., & Parasuraman, R. (1996). Psychophysiology and adaptive automation. *Biological Psychology, 42*(3), 249–268.

Chen, S., & Epps, J. (2013). Automatic classification of eye activity for cognitive load measurement with emotion interference. *Computer Methods and Programs in Biomedicine, 110*(2), 111–124.

Chen, S., Epps, J., & Chen, F. (2013). Automatic and continuous user task analysis using eye activity. In *Proceedings of the International Conference on Intelligent User Interfaces* (Santa Monica, CA) (pp. 57–66). New York, NY: ACM.

Chen, F., Ruiz, N., Choi, E., Epps, J., Khawaja, M. A., Taib, R.,… Wang, Y. (2012). Multimodal behaviour and interaction as indicators of cognitive load, *ACM Transactions on Intelligent Interactive Systems*, *2*(4), article 22.

Cohen, S., Kamarck, T., & Mermelstein, R. (1983). A global measure of perceived stress. *Journal of Health and Social Behavior, 24*(4), 385–396.

Conway, A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin and Review, 12*(5), 769–786.

Czerwinski, M., Horvitz, E., & Wilhite, S. (2004). A diary study of task switching and interruptions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 175–182). New York, NY: ACM.

de Greef, T., Lafeber, H., van Oostendorp, H., & Lindenberg, J. (2009). Eye movement as indicators of mental workload to trigger adaptive automation, in D. D. Schmorrow, I. V. Estabrooke and M. Grootjen (Eds.), *Foundations of augmented cognition. Neuroergonomics and operational neuroscience* (pp. 219–228). New York, NY: Springer.

De Waard, D. (1996). *The measurement of drivers' mental workload* (Unpublished doctoral thesis). Groningen University, Traffic Research Center, Groningen.

Gaillard, A. W. (1993). Comparing the concepts of mental load and stress. *Ergonomics, 36*, 991–1005.

Gaillard, A. W., & Wientjes, C. J. E. (1994). Mental load and work stress as two types of energy mobilization. *Work and Stress, 8*(2), 141–152.

Gevins, A., Smith, M. E., Leong, H., McEvoy, L., Whitfield, S., Du, R., & Rush, G. (1998). Monitoring working memory load during computer-based tasks with EEG pattern recognition methods. *Human Factors, 40*, 79–91.

Gopher, D., & Donchin, E. (1986). Workload—An examination of the concept. In K. R. Boff, L. Kaufman & K. P. Thomas, *Handbook of Perception and Human Performance* (pp. 41-1–41-49). New York, NY: John Wiley & Sons, Ltd.

Haapalainen, E., Kim, S., Forlizzi, J. F., & Dey, A. K. (2010). Psycho-physiological measures for assessing cognitive load. In *Proceedings of the International Conference on Ubiquitous Computing* (pp. 301–310). New York, NY: ACM.

Hancock, P. A., & Warm, J. S. (1989). A dynamic model of stress and sustained attention. *Human Factors, 31*(5), 519–537.

Hansen, J. H., & Patil, S. (2007). Speech under stress: Analysis, modeling and recognition. In *Speaker classification I* (pp. 108–137). Berlin: Springer, pp. 108–137.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In P. S. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139–183). Amsterdam: North-Holland.

Hendy, K. C., Hamilton, K. M., & Landry, L N. (1993). Measuring subjective workload: When is one scale better than many? *Human Factors, 35*(4), 579–601.

Hollender, N., Hofmann, C., Deneke, M., & Schmitz, B. (2010). Integrating cognitive load theory and concepts of human–computer interaction, *Computers in Human Behavior, 26*(6), 1278–1288.

Hollien, H. (1980). Vocal indicators of psychological stress. *Annals of the New York Academy of Sciences, 347*, 47–72.

Ikehara, C. S., & Crosby, M. E. (2005). Assessing cognitive load with physiological sensors. In *Proceedings of the Hawaii International Conference on System Sciences* (p. 295). Washington, DC: IEEE Computer Society.

Jarrold, C., & Towse, J. N. (2006). Individual differences in working memory. *Neuroscience, 139*(1), 39–50.

Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.

Kahneman, D., & Beatty, J. (1966). Pupil diameter and load on memory. *Science, 154*(3756), 1583–1585.

Klingner, J., Tversky, B., & Hanrahan, P. (2011). Effects of visual and verbal presentation on cognitive load in vigilance, memory, and arithmetic tasks. *Psychophysiology, 48*(3), 323–332.

Kramer, A. F. (1991). Physiological metrics of mental workload. In Damos, D. L. (Ed.), *Multiple task performance*. London: Taylor & Francis.

Larsen, R. J., & Fredrickson, B. L. (1999). Measurement issues in emotion research. In Kahneman, D., Diener E., & Schwarz N. (Eds.), *Well-being: The foundations of hedonic psychology* (pp. 40–60). New York, NY: Russell Sage Foundation.

Lin, Y., Zhang, W. J., & Watson, L. G. (2003). Using eye movement parameters for evaluating human–machine interface frameworks under normal control operation and fault detection situations. *International Journal of Human-Computer Studies, 59*(6), 837–873.

Makepeace, R., & Epps, J. (2015). Automatic task analysis based on head movement. In *37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 5167–5170). Piscataway, NJ: IEEE.

Martin, S. (2014). Measuring cognitive load and cognition: metrics for technology-enhanced learning. *Educational Research and Evaluation, 20*(7–8), 592–621.

McGrath, J. E. (1996). Stress and behavior in organizations. In Dunnette, M. D. (Ed.), *Handbook of industrial and organizational psychology*. Chicago, IL: Rand McNally.

Nygren, T. E. (1991). Psychometric properties of subjective workload measurement techniques: Implications for their use in the assessment of perceived mental workload. *Human Factors, 33*(1), 17–33.

O'Donnell, R. D., & Eggemeier, F. T. (1986). Workload assessment methodology. In K. R. Boff, L. I. Kaufman & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 1–49). New York, NY: John Wiley & Sons, Inc.

Oviatt, S. (2006). Human-centered design meets cognitive load theory: Designing interfaces that help people think. *Proceedings of the 14th ACM International Conference on Multimedia* (pp. 871–880). New York, NY: ACM.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. *Human Brain Mapping, 25*(1), 46–59.

Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurements as a means to advance cognitive load theory, *Educational Psychologist, 38*(1), 63–71.

Paas, F., Van Merriënboer, J. J. G., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills, 79*(1), 419–430.

Parasuraman, R., & Hancock, P. A. (2001). Adaptive control of mental workload. In P. A. Hancock & P. A. Desmond (Eds.), *Stress, workload, and fatigue. Human factors in transportation* (pp. 305–320). Mahwah, NJ: Lawrence Erlbaum Associates.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans, 30*(3), 286–297.

Peavler, W. S. (1974). Pupil size, information overload, and performance differences, *Psychophysiology, 11*(5), 559–566.

Pike, M., Wilson, M. L., Divoli, A., & Medelyan, A. (2012). "CUES: Cognitive Usability Evaluation System", in *Proceedings of the Second European Workshop on Human-Computer Interaction and Information Retrieval* (pp. 51–54). Aachen: CEUR Workshop Proceedings.

Reid, G. B., & Nygren, T. E. (1988). The subjective workload assessment technique: A scaling procedure for measuring mental workload, *Advances in Psychology, 52*, 85–218.

Reinerman-Jones, L. E., Matthews, G., Barber, D. J., & Abich, J. (2014). Psychophysiological metrics for workload are demand-sensitive but multifactorial. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (vol. 58, no. 1, pp. 974–978). Los Angeles, CA: Sage.

Rubio, S., Diaz, E., Martin, J., & Puente, J. M. (2004). Evaluation of subjective mental workload: A comparison of SWAT, NASA-TLX, and workload profile methods. *Applied Psychology, 53*(1), 61–86.

Ruiz, N. (2011). *Cognitive load measurement in multimodal interfaces* (Unpublished doctoral dissertation). University of New South Wales, Sydney, Australia.

Ruiz, N., Feng, Q. Q., Taib, R., Handke, T., & Chen, F. (2010). Cognitive skills learning: pen input patterns in computer-based athlete training. In *Proceedings of the International Conference on Multimodal Interfaces.* New York, NY: ACM.

Schuller, B., Steidl, S., Batliner, A., Epps, J., Eyben, F., Ringeval, F.,… Zhang, Y. (2014). The Interspeech 2014 computational paralinguistics challenge: Cognitive and physical load. In *Proceedings of the 15th Annual Conference of the International Speech Communication Association* (pp. 427–431). Baixas, France: International Speech Communication Association (ISCA).

Setz, C., Arnrich, B, & Tröster, G. (2010). Discriminating stress from cognitive load using a wearable EDA device. *IEEE Transactions on Information Technology in Biomedicine, 14*(2), 410–417.

Shen, J., Irvine, J., Bao, X., Goodman, M., Kolibaba, S., Tran, A.,… Dietterich, T. G. (2009). Detecting and correcting user activity switches: algorithms and interfaces. In *Proceedings of the 14th International Conference on Intelligent User Interfaces* (pp. 117–126). New York, NY: ACM.

Shi, Y., Ruiz, N., Taib, R., Choi, E., & Chen, F. (2007). Galvanic skin response (GSR) as an index of cognitive load. In *Proceedings of the International Conference on Human Factors in Computing Systems* (pp. 2651–2656). New York, NY: ACM.

Siegle, G. J., Ichikawa, N., & Steinhauer, S. (2008). Blink before and after you think: Blinks occur prior to and following cognitive load indexed by pupillary responses. *Psychophysiology, 45*, 679–687.

Siewiorek, D., Smailagic, A., & Starner, T. (2012). Wearable computers. In Jacko, J. (Ed.), *The Human-computer interaction handbook.* Boca Raton, FL: CRC Press.

Strayer, D. L., & William A. J. (2001). Driven to distraction: Dual-task studies of simulated driving and conversing on a cellular telephone. *Psychological Science, 12*(6), 462–466.

Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295–312.

Tepas, D. I., & Price, J. M. (2001). What is stress and what is fatigue? In P. A. Hancock & P. A. Desmond (Eds.), *Stress, workload and fatigue*, Mawah, NJ: L. Erlbaum.

Wilson, G. F., & Eggemeier, F. T. (1991). Performance-based and subjective assessment of workload. In D. L. Damos (Ed.), *Multiple task performance.* London: Taylor & Francis.

Yap, T. F., Epps, J., Ambikairajah, E., and Choi, E. H. C. (2015). Voice source under cognitive load: Effects and classification. *Speech Communication, 72*, 74–95.

Yeh, Y. Y., & Wickens, C. D. (1988). Dissociation of performance and subjective measures of workload, *Human Factors, 30*, 111–120.

Yin, B., Chen, F., Ruiz, N., & Ambikairajah, E. (2008). Speech-based cognitive load monitoring system. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing* (pp. 2041–2044). Piscataway, NJ: IEEE.

Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: Mental workload in ergonomics. *Ergonomics, 58*(1), 1–17.

Yu, K., Epps, J., & Chen, F. (2011). Cognitive load evaluation of handwriting using stroke-level features. In *Proceedings of the International Conference on Intelligent User Interfaces* (pp. 423–426). New York, NY: ACM.

Zarjam, P., Epps, J., Chen, F., & Lovell, N. (2013). Estimating cognitive workload using wavelet entropy-based features during an arithmetic task, *Computers in Biology and Medicine, 43*(12), 2186–2195.

Zarjam, P., Epps, J., and Lovell, N. (2015). Beyond subjective self-rating: EEG signal classification of cognitive workload. *IEEE Transactions in Autonomous Mental Development, 7*(4), 301–310.

12

Comparing Mobile Experience

Xiaoge Xu

Introduction

Mobile has become more powerful than any other medium, enabling us to do almost anything, anytime, anywhere. With mobile, we can do shopping without going to a shopping mall, transfer money instead of going to a bank, see a doctor without going to a hospital, learn anytime, anywhere instead of going to a school, or work without going to a work place, to name just a few examples. Indeed, mobile has been changing the way we communicate, work, and live.

Our increasing and widening use of mobile has brought us a ubiquitous and unique mobile experience. As a working definition, mobile experience can be defined as both a process and an outcome of a user's interaction with a product, a service, a content, or their combination, in an interactive, personalized, immersive, and mobile context. Mobile experience can fall into a wide spectrum ranging from the worst to the best.

Derived from user experience, mobile experience has been examined by previous scholars within the overall framework of user experience (e.g. Eune & Lee, 2009; Moor et al., 2010). This framework, however, has different components proposed by different scholars. Generally, there are two broadly defined categories. The first category may consist of the user's awareness, adoption, learning, and use (Alben 1996; Norman 1999, 2004) whereas the second can cover anticipation, encounter, interpretation, judgment, internalization, and recount of experience (McCarthy & Wright, 2007).

Within the user experience framework, what distinguishes mobile experience from user experience is its ubiquitous and unique interaction, personalization, and immersion. In other words, a mobile user's interaction with and personalization of a product, a service, content, or their combination can occur anytime anywhere. Being immersive, mobile experience enables a user to be immersed in a mixed reality, reshaping time, space, and dimension.

Due to diversified political, social, economic, and cultural differences on the macro level, and differences in mobile uses, processes and effects on the micro level in different countries, mobile experience differs from country to country. Differences can be normative, meaning that mobile experience can be different in terms of what

The Wiley Handbook of Human Computer Interaction, Volume 1, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

users expect to have. Differences can also be empirical in terms of what they actually obtain. Furthermore, differences also lie in the gap between normative and empirical mobile experience.

All the differences mentioned above, however, cannot be effectively mapped, measured, and modeled without comparing it. The rising need to compare mobile experience in different countries is also largely due to the fact that mobile experience can be shaped and reshaped differently by cultural, economic, social, and political differences in different countries. Furthermore, different countries around the world have become so interdependent and interconnected in so many ways that whatever we share or differ will affect our increasing interaction with each other in the mobile space. Therefore, it is imperative for us to compare mobile experience so that we can have a better understanding and command of it to enhance our mobile media, communication, and services in this increasingly interdependent and interconnected world.

A Six-stage Approach

Earlier studies on mobile experience fall largely into three major approaches: (a) culture focused, (b) technology centric, and (c) user oriented. Taking the culture-focused approach, earlier studies mainly investigated "the distinctive cultural inclinations of each country in user experiences related to mobile phones" and mapped "cultural models to mobile phone user interface design for applying to design practice" (Eune & Lee, 2009). While technology-centric studies investigated, for instance, "functionality, usability, and experience" in a combined fashion (McNamara & Kirakowski, 2005), user-oriented studies examined usefulness, ease of use, hedonics, aesthetics, and pleasure/fun (Mahlke, 2005) on the one hand and "the challenge of continuously involving the user and the need for tools to facilitate the integration of knowledge into the increasingly interdisciplinary development process" on the other hand (Moor et al., 2010).

Taking a culture-focused, technology-centric, or user-oriented approach, few previous studies have managed to provide a holistic picture of mobile experience since they would neglect or belittle other elements when they focused on one particular element. The neglected or belittled elements, however, might otherwise be important in examining mobile experience. Another disturbing factor is that few studies have fully or successfully examined mobile experience at the normative level and the empirical level, and the gap between the two. As a result, similarities and differences in mobile experience have not been fully or successfully mapped, measured, or modeled.

As a remedy to this dilemma, a holistic approach can be employed in comparing mobile experience. And this approach involves the following six different stages: (a) enticement, (b) entertainment, (c) engagement, (d) empowerment, (e) enlightenment, and (f) enhancement. Theoretically, this six-stage approach is based on a combination of the widely used Maslow's hierarchy of needs (1943, 1954), its amended version (Koltko-Rivera, 2006), and the hierarchy of desires (Turner, 2003).

Based on his motivation-oriented needs, Maslow (1943, 1954) proposed a hierarchy of five needs: physiological needs (such as air, food, drink, shelter, warmth, sex, and sleep), safety needs (such as protection from elements, security, order, law, stability, freedom from fear), love and belonging needs (such as friendship, intimacy, affection and love in human social interactions such as family, friends, officemates, romantic relationships), esteem needs (such as self-esteem, achievement, mastery, independence, status, dominance, prestige, and managerial responsibility), and self-actualization (such as realizing personal potential, self-fulfillment, and seeking personal growth) (McLeod, 2007). These different needs, according to Maslow, are to be met in their hierarchical sequence. In other words, only after the physiological need, the very first fundamental need, is fulfilled, will a person move on to fulfill the second need, i.e. safety, and so on.

To improve his five-stage model, Maslow (1970a, 1970b) added three additional ones, that is, cognitive needs (such as knowledge and meaning), aesthetic needs (such as appreciation and search for beauty, balance, form), and transcendence needs (such as helping others to achieve self-actualization). The extended model is also sequential in that after the first original four needs, he added cognitive needs and aesthetic needs. After self-actualization, the eight-stage model ends with transcendence needs (Koltko-Rivera, 2006; McLeod, 2014). Although his model has been improved in capturing more stages of human needs, the same limitation remains unchanged: the stages remain sequential, which may not be necessarily so in terms of human needs to be met in realty.

Unlike Maslow's sequence-centric stage approach, the hierarchy of desires emphasizes the strength of layered desires (primary and second-order desires), although it also sticks with the hierarchy (Turner, 2003). The difference, however, lies in the fact that the sequence can be changed by the change in importance that a person attaches to his or her desires.

Equipped with the theoretical frameworks cited above, the six-stage approach states that, although the six-stage approach can be as sequential as Maslow's five- or eight-stage approach, its focus lies more in the sequence of importance instead of exposure because the sequence of importance can change without necessarily following the sequence of exposure. In other words, if a mobile user is an interactivity person, he or she may attach greater importance to engagement, which can be viewed as the most important stage.

Being enticed into a product, a service, a content or their combination is the very first essential step for us to walk into the mobile space. How enticed are we is a crucial element that determines our decision whether we should consume a product, a service, content, or their combination. Without being enticed, it would be hard for us to stay in the mobile space and enter other aspects of mobile experience. It is especially true in the mobile world where our attention and interest are swiftly moving around in a highly mobile context.

After being enticed, we expect to be entertained in one way or another to a different degree, depending on the nature of what we consume, be it a product, a service, content or their combination. Being entertained means a lot in maintaining our attention and interest in our consumption process. It has been largely neglected or even ignored in a product, a service, a content or their combination of less entertaining nature. Being less entertaining does not necessarily mean it can be neglected or ignored. This is especially true among entertainment-oriented mobile users.

Being enticed and entertained, we also expect to be engaged all the time without roaming away from whatever we are supposed to consume. In the mobile space, almost everything seems to be so enticing and interesting that we can be easily tempted to swipe away from one product to another, from one service to another, from one content to another, or from one combination to another. How to keep us fully engaged all the time constitutes a big challenge to mobile product, service, or content providers.

Besides enticement, entertainment, and engagement, it is also our strong desire and expectation to be empowered in consuming a product, a service, content, or their combination. To empower us instead of forcing us is a big and effective factor that distinguishes the best from the crowd in terms of how to consume a product, a service, content, or their combination.

To be enlightened can be viewed as one of the highest levels of mobile experience. In other words, after consuming a product, a service, content or their combination, we have obtained a better awareness and understanding of what is being consumed or everything related. And we have also increased our knowledge.

The ultimate highest level of mobile experience is to enable us to experience enhanced interaction with a product, a service, a content or their combination through enhancing our consciousness, skills, and abilities.

A 3M Process

Equipped with the proposed six-stage approach to mobile experience, we can compare mobile experience through a 3M process: mapping, measuring, and modeling. Through mapping, we can locate where similarities and differences in mobile experience lie. After locating where they are, the next step is to measure to what extent mobile experience is similar or different. The third step is to model mobile experience. In other words, the third step is designed to provide an explanation and even prediction of mobile experience. Both mapping and measuring can occur at six different stages. Mapping and measuring can also happen on both the normative level and the empirical level. A gap between the normative and the empirical levels will be identified and measured after mapping and measuring the normative and empirical mobile experience. Furthermore, our comparison can be conducted on the micro level, meaning we can conduct our three-dimension comparison at each of the six stages. Our macro comparison can also be threefold by mapping and measuring the total mobile experience from the first stage through the last stage. After mapping and measuring mobile experience, we propose a model to describe, explain, and predict changes and patterns in mobile experience around the world.

Mapping mobile experience

Before comparing mobile experience, it is essential to map where it is located. As mobile experience is situated at six stages, it is natural to map it through looking into specific components of mobile experience at each stage.

At the enticement stage, we look into the following three fundamental dimensions of mobile experience: (a) appealing interface, (b) easy navigating, and (c) interest arousing. By appealing interface, we do not mean any fancy or sophisticated interface. Instead, we refer to simple, easy, and attractive interface, which can attract and maintain our attention and interest. Easy navigation means a simple and easy device-specific navigation mechanism, allowing us to navigate easily from one place to another on the screen. By "interest arousing" we mean the home interface should arouse our interest in reading or viewing a product, a service, content, or their combination anytime, anywhere.

Mobile experience at the entertainment stage can be located in the following areas: (a) fun, (b) pleasure, and (c) satisfaction in using a product, a service, content or their combination. Being fun refers to what we consume being amusing, entertaining, and enjoyable while pleasure means sensual gratification. Both fun and pleasure are related to methods or delivery. But fun focuses on the emotional response while pleasure highlights the sensual part. Satisfaction refers to fulfillment of our wishes, expectations, or needs.

At the engagement stage, we map mobile experience through (a) searching, (b) interacting, and (c) sharing. To search before, in the middle of, or even after consuming a product, a service, content or their combination is to keep us engaged in our consumption. Our engagement also involves interacting with others via social media or online chatting or interacting mechanisms such as chatrooms, blogging, emailing, and instant messengers. To share with whatever we have consumed also enables us to stay engaged during our consumption.

Mobile experience can also be located at the empowerment stage by looking into the following: (a) selecting (content, font size, layout, color, background, theme, region, media, edition, and language), (b) commenting (correcting, rating, and feedback), and (c) producing (tweeting, blogging, writing, editing, publishing, and broadcasting). Mobile experience at this stage largely refers to the extent we are empowered to select, comment, and produce whatever we would like via mobile and the corresponding outcomes.

Enlightenment-specific mobile experience can be located in the following aspects: (a) awareness, (b) understanding, and (c) consciousness. To consume a product, a service, content or their combination will enable us to be enlightened in one way or another through enhancing our awareness, understanding, and consciousness, respectively, of a product, a service, or their combination.

Ultimately, mobile experience can be mapped through the following ways: (a) knowledge, (b) skills, and (c) abilities. The ultimate goal of enriching mobile experience is to enhance our interaction with a product, a service, a content or their combination through enhancing our knowledge, skills, and abilities respectively.

After mapping mobile experience comprehensively, we will have a better and holistic picture of where mobile experience is specifically situated and where similarities and differences are located before we can start to measure mobile experience.

Measuring mobile experience

To measure mobile experience, previous studies have employed different methods. Among them are (a) use of metrics to describe and measure it, (b) using stories to provide its rich details, (c) using usability tests to gauge it, and (d) using a hierarchy of user needs to measure its different layers.

For example, in using metrics, some of earlier studies employed a number of metrics to describe and measure user experience (Hassenzahl, 2003; Jordan, 2000; Norman, 2004).

The story approach can be illustrated by other studies, which provided stories of how users experience a product or service (Forlizzi & Ford, 2000) or still other studies, which provided in-depth views and insights or social and cultural factors located in narratives or storytelling in the hope of capturing and interpreting user experience in a holistic and constructionist way by listening to users' stories of experience with technology (McCarthy & Wright 2005).

The usability approach can be exemplified in other earlier studies, where scholars focused on either "the ability to complete some functional or goal-directed task within a reasonable time" or "the degree, to which a product is desirable or serves a need beyond the traditional functional objective" (Logan, 1994). Some of earlier usability studies also examined the following new dimensions: (a) "efficiency, affect, control, and learnability, and helpfulness" (Kirakowski & Corbett, 1993), (b) "performance and image/impression" (Han, Yun, Kwahk, & Hong, 2001), and (c) "ease of use, helpfulness, affective aspect, minimal memory load, and efficiency" (Ryu & Smith-Jackson, 2006).

Finally, the hierarchy approach can be illustrated by a study that investigated a hierarchy of user needs: safety, functionality, usability, and a pleasurable experience (Jordan, 2000).

To measure either normative or empirical mobile experience, a combination of surveys, focus groups, interviews, and a comparative content analysis can be used. To standardize the quantification of both normative and empirical mobile experience, we use a scale of 1–5 points, with 1 being the least and 5 being the most, to measure each of the three components at each of the six stages of mobile experience. Both normative and empirical mobile experience can also be measured on both micro and macro levels. By micro level, we refer to the level of mobile experience at each stage of mobile experience while the macro level is meant to be about the total level of all six stages of mobile experience.

On the micro level, each component of mobile experience can be measured in three ways: (a) the normative measurement of each component of mobile experience, (b) the empirical measurement of each component of mobile experience, and (c) the gap measurement between the normative and empirical level of each component of mobile experience. At each stage, at both the normative and empirical levels, a scale of 1–5 is assigned to each component.

On the macro level, mobile experience can be measured in three ways: (a) measuring the total normative mobile experience, (b) measuring the total empirical mobile experience, and (c) measuring the gap between the normative and empirical measurement of the total mobile experience. The mobile experience gap can be gauged by using the following formula: the gap = the total normative mobile experience (90) – the total empirical mobile experience (90). The gap may fall under any of the three categories: the narrowest gap (1–30), the medium gap (31–60) and the widest gap (61–90) (see Table 12.1).

Modeling mobile experience

After mapping and measuring these differences, the ultimate goal of our investigation is to locate different factors that have shaped the differences in normative mobile experience, empirical mobile experience, and the gap between the two.

Earlier studies have identified various factors that have shaped or influenced mobile experience in terms of dimensions, density, and direction. And these factors can be

Table 12.1 Measuring mobile experience: Indicators and measurements.

Stages	*Indicators*	*Measurement*	*Score*
Enticed	Appealing interface	1 least … 5 most	
	Easy navigation	1 least … 5 most	
	Interest arousing	1 least … 5 most	
		Subtotal	
Entertained	Fun	1 least … 5 most	
	Pleasure	1 least … 5 most	
	Satisfaction	1 least … 5 most	
		Subtotal	
Engaged	Searching	1 least … 5 most	
	Interacting	1 least … 5 most	
	Sharing	1 least … 5 most	
		Subtotal	
Empowered	Selecting	1 least … 5 most	
	Commenting	1 least … 5 most	
	Producing	1 least … 5 most	
		Subtotal	
Enlightened	Awareness	1 least … 5 most	
	Understanding	1 least … 5 most	
	Consciousness	1 least … 5 most	
		Subtotal	
Enhanced	Knowledge	1 least … 5 most	
	Skills	1 least … 5 most	
	Abilities	1 least … 5 most	
		Subtotal	
		Grand total	

grouped into two categories: internal and external factors (see Table 12.2). The internal factors refer to those that are closely related to all dimensions of mobile such as users, devices, apps, and networks while the external factors are the social, cultural, and contextual elements that may shape differences in mobile experience.

Most earlier studies focused on factors that have shaped empirical experience, leaving the normative mobile experience and the gap between the normative and empirical mobile experience largely untouched. Even within the examination of the empirical mobile experience, earlier studies did not investigate whether factors such as connection, curiosity, consumption, competiveness, and creativity, which may have a role to play in shaping mobile experience. Connection refers to mobile connection, which constitutes the foundation of a mobile space, without which everything mobile is impossible. As a shaping factor, curiosity can enable mobile users to stay interested in exploring everything mobile. The higher level of curiosity a mobile user has, the higher level of his or her desire to experience everything mobile. Consumption refers to the overall consumption spending of a country. The level of consumption spending may influence the level of mobile experience too at different stages. Competitiveness refers to the level of the overall competitiveness of a country, which can also influence the level of mobile experience. Creativity refers to the level of being creative of a nation, which can also be a major shaping factor in influencing mobile experience.

Table 12.2 Factors shaping mobile experience: Internal and external.

Internal	*Factors*	*Studies*
	Location, social context, mobility, battery, application interface design, application performance, phone features, routine, and lifestyle needs	Ickin et al. (2012)
	emotions and memories	Kujala and Miron-Shatz (2013)
	ease of use, awareness, security, usefulness, and availability/accessibility,	Sarmento and Patrício (2012)
	mobile site optimization,	Djamasbi et al. (2014)
	motor impairments and accessibility,	Naftali and Findlater (2014)
	gender differences such as responsive mobile design for female while dedicated mobile design for male,	Silverstein (2014)
	trust and privacy	Krontiris, Langheinrich, & Shilton (2014)
	Load speed, site format, calculated download speed, social media presence, and app presence	Silverstein (2013a, b)
	personalization of content, structural navigation and representation	Chua, Wan, Chang, and Yi (2014)
	device capabilities and settings, network performance	Patro, Rayanchu, Griepentrog, Ma, & Banerjee (2013)
	values, emotions, expectations, prior experiences, physical characteristics, personality, motivation, skills, usefulness, reputation, adaptivity	Arhippainen and Tähti (2003)
	attitude, social influence, media influence, perceived mobility, and perceived monetary value	Hong, Thong, Moon, and Tam (2008)
External	*Factors*	*Studies*
	Social structure such as degree of homogeneity, extent to which egalitarian, communication patterns with the outside world, gender, ethnicity, generational cultures, religion, education provision, and support for literacy and language, temporal structures of daily life (subjective experience of time, societal time use structures and cultural expectations about time), values (openness to technological innovation, the degree to which societies are individualistic or group oriented, other culture values), communication (communication patterns and expectations and low and high context cultures), material cultures (special considerations, national differences in housing characteristics, and artefacts)	Thomas, Haddon, Gilligan, Heinzmann, and de Gournay (2005)
	social factors: time pressure, pressure of success and failure, etc. Cultural factors: habits, norms, religion, etc. Context of use: time, place, accompanying persons, temperature, etc.	Arhippainen and Marika Tähti (2003).

After identifying shaping factors, it is crucial to design a model to describe, explain, and predict mobile experience. Among previous model-building efforts are the following five models proposed by Pine and Gilmore (1999): (a) performance model, which metaphorically allows actors to interact with guests on a stage in a personal way; (b) four realms of an experience, which consists of entertainment, education, escapism, and esthetics; (c) customization, which enables app developers and mobile users transparently collaborate to determine and satisfy users' needs; (d) progression of economic value model, in which guests respond to the perceived value and obtain rich and effective educational experiences; (e) four S model, in which guests experience from satisfaction, through sacrifice and surprise, to suspension, with the basic principle to lead to high levels of fulfillment and loyalty among guests, and (f) other models including the socio-culture model, the design constraint model, the leisure paradox model, and the service triangle model (cited in Sims, Williams & Elliot, 2007).

Previous models can hardly describe, explain, and predict mobile experience because they omit or ignore some basic components or stages of mobile experience. To remedy the situation, Sims, Williams, and Elliot (2007) proposed a cognitive model, which creates and manages representations that inhabit an agent's inner world, namely sensations, perceptions, conceptions / knowledge and simulations. Although effective in measuring the quality of experience, their model, like others, fails to capture experience at different stages on different levels. In other words, previous models fail to describe, explain, and predict the normative and empirical level of mobile experience as well as the gap between the two at six stages of mobile experience.

To fill the gap, a factor-stage-dimension framework is proposed. It consists of five factors (connection, curiosity, consumption, competiveness, and creativity) that influence six stages of mobile experience (enticed, entertained, engaged, empowered, enlightened, and enhanced) in three dimensions (the normative, the gap, and the empirical dimensions) (see Table 12.3).

In this framework, connection refers to mobile connection, which is fundamental and essential as there would be no mobile communications of any kind if there was no mobile connection. If there was no mobile communication, there would be no mobile experience at all. Furthermore, the extent of connection also determines the extent of mobile experience. Simply put, more connections mean mobile experience being enjoyed by more mobile users. This important factor has not been fully examined in terms of how it may influence or affect the level of mobile experience on the normative level, the empirical level and the gap between the two. The first hypothesis is that the level of mobile connection in a country is correlated with its level of mobile experience.

Table 12.3 Mobile experience: Five factors, six stages, and three dimensions.

Five factors	*Six stages*	*Three dimensions*
Connection	Enticed	
Curiosity	Entertained	Normative
Consumption	Engaged	Gap
Competitiveness	Empowered	Empirical
Creativity	Enlightened	
	Enhanced	

Equally essential to mobile experience is the level of curiosity in a country. If a nation is generally not curious or its level of curiosity is low, not matter how well a country is connected, it won't be able to enjoy mobile experience to the full. It is well known that many mobile features and functions remain unknown to most mobile users, not to mention fully used by them. Beyond mobile, there are so many unknown things to be explored and experienced. It is curiosity that leads human beings to the unknown world. This factor has not yet fully been investigated in connection with mobile experience. So it is hypothesized that the level of curiosity in a country is associated with its level of mobile experience.

Consumption refers to the volume of using mobile including using a product, a service, a content or their combination in a country. The more a country uses a product, a service, a content or their combination, the more experience it will be exposed to or obtain. It is hypothesized that the level of a country's consumption can be correlated with its level of mobile experience.

Staying competitive is another influencing factor, which has also been largely belittled if not neglected totally. The more competitive a nation is, the more experience will be enriched and enlarged, which will be enjoyed by more of its mobile users. It is therefore also hypothesized that the level of competitiveness of a nation is correlated with its level of mobile experience.

Creativity is the last but definitely not the least factor. As a matter of fact, it is far more powerful in influencing the level of mobile experience in a country. The more creative a nation is in general, the more it will expect from mobile experience at different stages. It is therefore hypothesized that the higher level of creativity of a country leads to the higher level of mobile experience in that country.

The five factors are combined to constitute the environment that can shape different stages of mobile experience. The level of each of the five factors also constitutes the individual level of the environmental conduciveness. Combined, the five levels represent the level of the total environmental conduciveness. How the environmental conduciveness is correlated with mobile experience is presented visually in Figure 12.1.

Based on the factor-stage-dimension framework and the overall correlation between the environmental conduciveness and mobile experience, the following three hypotheses are proposed. End each hypothesis has its own respective five subhypotheses.

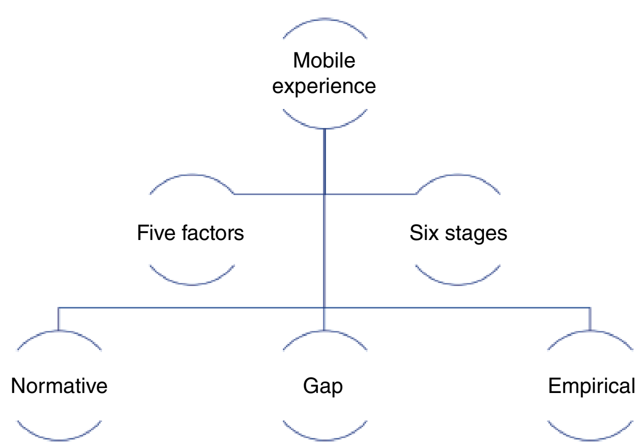


Figure 12.1 Interconnections between mobile experience and its related elements.

- H1. Environmental conduciveness is correlated with normative mobile experience.
  - H1a: Mobile connection is correlated with normative mobile experience.
  - H1b: User curiosity is correlated with normative mobile experience.
  - H1c: User consumption is correlated with normative mobile experience.
  - H1d: User competitiveness is correlated with normative mobile experience.
  - H1e: User creativity is correlated with normative mobile experience.
- H2. Environmental conduciveness is associated with empirical mobile experience.
  - H2a: Mobile connection is associated with empirical mobile experience.
  - H2b: User curiosity is associated with empirical mobile experience.
  - H2c: User consumption is associated with empirical mobile experience.
  - H2d: User competitiveness is associated with empirical mobile experience.
  - H2e: User creativity is associated with empirical mobile experience.
- H3. Environmental conduciveness is correlated with the gap between its normative and empirical mobile experience.
  - H3a: Mobile connection is correlated with the gap between its normative and empirical mobile experience.
  - H3b: User curiosity is correlated with the gap between its normative and empirical mobile experience.
  - H3c: User consumption is correlated with the gap between its normative and empirical mobile experience.
  - H3d: User competitiveness is correlated with the gap between its normative and empirical mobile experience.
  - H3e: User creativity is correlated with the gap between its normative and empirical mobile experience.

To place all these hypotheses in perspective, the following diagram shows their interconnections, interactions, and interinfluences, which can serve as a model to describe, explain, and predict mobile experience in a country, and can be used as a model to compare mobile experience in different countries around the world (see Figure 12.2).

Concluding Remarks

In this increasingly interconnected and interdependent world, it is imperative to compare mobile experience in different countries for better cross-country mobile communications in different areas. Limited by culture-focused, technology-centric, or user-oriented approaches, earlier studies failed to provide a comprehensive comparison of mobile experience. In this chapter, a six-stage (enticed, entertained, engaged, empowered, enlightened, and enhanced) approach was proposed to offer stage-specific comparison of different components of mobile experience. As far as the comparison process is concerned, few studies have located, gauged, and explained mobile experience in a combined investigation. To fill the void, a 3M (mapping, measuring, and modeling) process was suggested with a special focus on the gap between the normative and the empirical dimensions of mobile experience. Combining the six-stage approach, the 3M process and the gap focus, a factor-stage-dimension framework was recommended to provide an overall guidance for generating some possible patterns to explain and predict changes and trends in mobile experience.

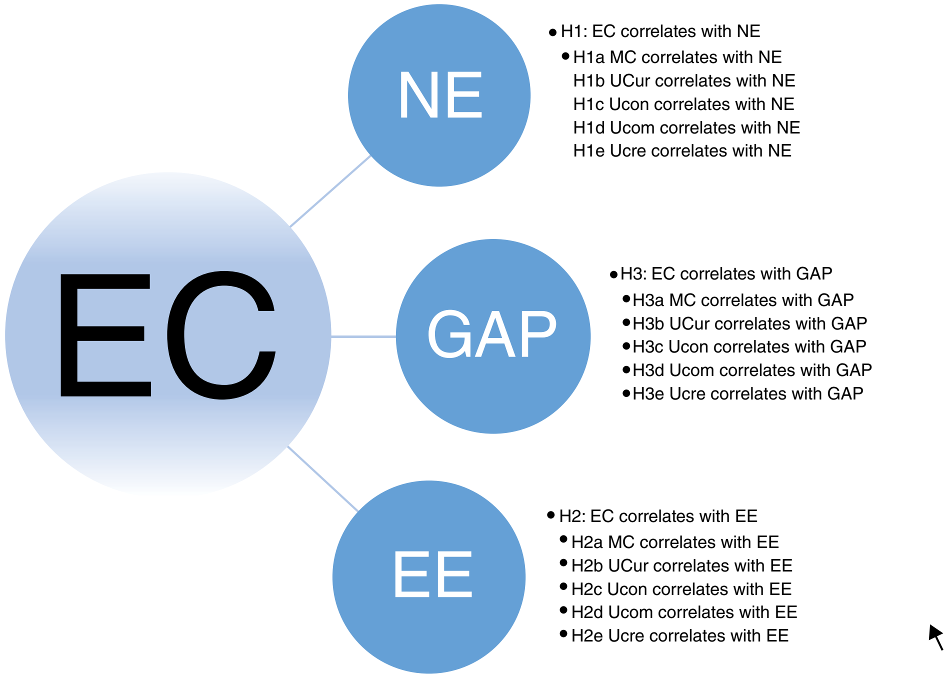


Figure 12.2 Three hypotheses and their respective subhypotheses.

References

Alben, L. (1996). Quality of experience: Defining the criteria for effective interaction design. *Interactions 3*(3), 11–15.

Chua, W. Y., Wan, M. P. H., Chang, K., & Yi, W. (2014, August). *Improving mobile applications usage experience of novice users through user-acclimatized interaction: A case study.* Paper presented at the 20th Americas Conference on Information Systems (AMCIS 2014), Savannah, GA.

Djamasbi, S., McAuliffe, D., Gomez, W., Kardzhaliyski, G., Liu, W., & Oglesby, F. (2014, June). Designing for success: Creating business value with mobile user experience (UX). In *International Conference on HCI in Business* (pp. 299–306). Cham: Springer.

Eune, J., & Lee, K. P. (2009). Analysis on intercultural differences through user experiences of mobile phone for globalization. In *Proceedings of the International Association of Societies of Design Research* (pp. 1215–1226). n.p.: IASDR.

Forlizzi, J., & Ford, S. (2000, August). The building blocks of experience: An early framework for interaction designers. In *Proceedings of the Third Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques* (pp. 419–423). New York, NY: ACM.

Han, S. H., Yun, M. H., Kwahk, J., & Hong, S. W. (2001). Usability of consumer electronic products. *International Journal of Industrial Ergonomics, 28*, 143–151.

Hassenzahl, M. (2003). The thing and I: Understanding the relationship between user and product. In M.A. Blythe, A. F. Monk, K. Overbeeke & P. C. Wright (Eds.), *Funology: From usability to enjoyment* (pp. 31–42). Norwell, MA: Kluwer Academic Publisher.

Ickin, S., Wac, K., Fiedler, M., Janowski, L., Hong, J. H., & Dey, A. K. (2012). Factors influencing quality of experience of commonly used mobile applications. *Communications Magazine, IEEE, 50*(4), 48–56.

Jordan, P. W. (2000). *Designing pleasurable products.* New York, NY: Taylor & Francis.
Kirakowski, J., & Corbett, M. (1993). SUMI: The software usability measurement inventory. *British Journal of Educational Technology, 24*(3), 210–212.
Koltko-Rivera, M. E. (2006). Rediscovering the later version of Maslow's hierarchy of needs: Self-transcendence and opportunities for theory, research, and unification. *Review of General Psychology, 10*(4), 302–317.
Krontiris, I., Langheinrich, M., & Shilton, K. (2014). Trust and privacy in mobile experience sharing: Future challenges and avenues for research. *IEEE Communications Magazine, 52*(8), 50–55.
Logan, R. J. (1994). Behavioral and emotional usability: Thomson Consumer Electronics. In M. E. Wiklund (Ed.), *Usability in practice* (pp. 59–82). New York, NY: AP Professional.
Mahlke, S. (2005). Understanding users' experience of interaction. In *Proceedings of the 2005 Annual Conference on European Association of Cognitive Ergonomics* (pp. 251–254). Chania, Greece: University of Athens.
Maslow, A. H. (1943). A theory of human motivation. *Psychological Review, 50(4)*, 370–396.
Maslow, A. H. (1954). *Motivation and personality.* New York, NY: Harper & Row.
Maslow, A. H. (1970a). *Motivation and personality.* New York, NY: Harper & Row.
Maslow, A. H. (1970b). *Religions, values, and peak experiences.* New York, NY: Penguin.
McCarthy, J., & Wright, P. C. (2005). Putting "felt-life" at the centre of human-computer interaction (HCI). *Cognition, Technology, and Work, 7*, 262–271.
McLeod, S. A. (2007). *Maslow's hierarchy of needs.* Retrieved from www.simplypsychology.org/maslow.html
McNamara, N., & Kirakowski, J. (2005). Defining usability: quality of use or quality of experience? In *Proceedings of the International Professional Communication Conference* (pp. 200–204). Piscataway, NJ: IEEE.
Moor, K. D., Berte, K., Marez, L. D., Joseph, W., Deryckere, T., & Martens, L. (2010). User-driven innovation? Challenges of user involvement in future technology analysis. *Science and Public Policy, 37*(1), 51–61.
Naftali, M., & Findlater, L. (2014). Accessibility in context: Understanding the truly mobile experience of smartphone users with motor impairments. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 209–216). New York, NY: ACM. Retrieved from http://msx.sagepub.com/content/18/3/275.short
Norman, D. (1999). *The invisible computer.* Cambridge, MA: MIT Press.
Norman, D. (2004). *Emotional design: Why we love (or hate) everyday things.* New York, NY: Basic Books.
Patro, A., Rayanchu, S., Griepentrog, M., Ma, Y., & Banerjee, S. (2013). Capturing mobile experience in the wild: A tale of two apps. In *Proceedings of the Ninth ACM Conference on Emerging Networking Experiments and Technologies* (pp. 199–210). New York, NY: ACM.
Pine, B. J., & Gilmore, J. H. (1999). *The experience economy: Work is theatre and every business a stage.* Boston, MA: Harvard Business Press.
Ryu, Y. S., & Smith-Jackson, T. L. (2006). Reliability and validity of the mobile phone usability questionnaire (MPUQ). *Journal of Usability Studies, 2*(1), 39–53.
Sarmento, T., & Patrício, L. (2012). *Mobile service experience—a quantitative study.* Retrieved from http://comum.rcaap.pt/handle/10400.26/7552
Sims, F., Williams, M. A., & Elliot, S. (2007, July). Understanding the mobile experience economy: A key to richer more effective M-business technologies, models and strategies. In *International Conference on the Management of Mobile Business* (p. 12). Piscataway, NJ: IEEE.
Silverstein, M. (2013a, September). Introducing the mobile experience. *Scorecard*. Retrieved from http://www.thesearchagents.com/2013/09/introducing-the-mobile-experience-scorecard/

Silverstein, M. (2013b September). The Search Agency's mobile experience. *Scorecard: Fortune 100 Companies.* Retrieved from http://www.thesearchagents.com/2013/10/the-search-agencys-mobile-experience-scorecard-fortune-100-companies/

Silverstein, M. (2014). Female-oriented sites more optimized for mobile? [Web log post]. Retrieved from http://www.thesearchagents.com/2014/03/female-oriented-sites-more-optimized-for-mobile/

Turner, R. (2003). *Desire and the self* (research paper). Retrieved from http://robin.bilkent.edu.tr/desire/desire3.pdf

13

Factors of Immersion

Noirin Curran

What is Immersion?

Herodotus, a Greek historian from the fifth century BC, wrote a story about the origin of games and their invention by the ancient Lydians, who had suffered a terrible famine (Rawlinson, 1861, p. 182). In order to help combat hunger, they planned that the citizens would eat during the first day, and the next day they would become so deeply engaged in dice games, so immersed, that they would not feel hunger and would not need to eat. They planned to continue in this pattern until the famine ended. Even in this story of the origin of dice games, there is a mention of players being so engaged in a game that they almost forget themselves, and not only that, but immersion is seen as being the aim of the game.

Jennett et al. speak for many when they characterize immersion as being "key to a good gaming experience" (2008, p. 644). Despite there being, as yet, no single accepted definition of the concept, it is now commonplace to describe digital games in terms of being *immersive*.

Brown and Cairns (2004) define immersion, simply, as "the degree of involvement with a game" (p. 1298). In their description of immersion, Coomans & Timmermans (1997, p. 6) outline immersion as being a feeling of being deeply engaged where people "enter a make-believe world as if it is real." While these are broad definitions, they are acceptable, and because of their nature they are more universally applicable. Some broad definitions of immersion refer to experiencing a level of virtual engagement or to cutting off or disassociation from reality and entering a virtual world (Brown & Cairns, 2004; Coomans & Timmermans, 1997; Haywood & Cairns, 2005; Jennett et al., 2008) while other definitions refer to an immersive experience as substituting virtuality for reality for a period of time or becoming physically part of the experience (Grimshaw, 2007; Pine & Gilmore, 1999). The characteristics of being immersed are considered to be, among others, a lack of awareness of time and of the real world around you, and a sense of actually being present within the task environment—immersion gives a sense of real-world dissociation (Haywood & Cairns, 2005; Jennett et al., 2008).

Immersion is becoming increasingly important as a hitherto neglected part of human computer interaction, and although it is most often mentioned in the context

The Wiley Handbook of Human Computer Interaction, Volume 1, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

of games, it can also have significance when reading a particularly engrossing book or watching a movie or television show in which the individual identifies with a character (McCarthy & Wright, 2004; Schubert & Crusius, 2002). "When identifying with a character in a book or movie, individuals tend to put themselves in the character's place, and in a sense, experience what that character experiences" (Witmer & Singer, 1998, p. 227). The experience of being immersed in a book or a film is significantly different to being immersed in a game, however. In most media, the character with which the player identifies will follow the plot regardless of outside input, as the role is already predetermined by the writer or director. In contrast, in a game, the participant has control over the character and the game actions, and this sense of agency or ability to impact the game environment is an important element of immersion (McMahan, 2003; Murray, 1997).

Immersion, as a concept, also fits within a body of research on equally fascinating levels of involvement such as flow, engagement, presence, and cognitive absorption.

Flow is a concept developed by Csíkszentmihályi (1990), which encompasses online applications as well as sports and artistic pursuits and involves emotions of enjoyment caused by a sense of "balance between skill and challenge…in the process of performing an intrinsically rewarding activity" (Brockmyer et al., 2009, p. 625). It is described as being "an optimal and therefore extreme experience" (Jennett et al., 2008 p. 642), and inherently pleasurable (Murray, 1997, p. 98), while immersion is not always an extreme experience and is not always pleasurable.

Engagement is described as "something that draws us in, that attracts and holds our attention" (Chapman, 1997, p. 3) and "a desirable even essential human response to computer-mediated activities" (Laurel, 1993, p. 112). When compared with immersion, engagement is described as "a more deliberate, critical mode of participation" (Carr, Buckingham, Burn, & Schott, 2006, p. 54). Although it is not unique to games, McMahan (2003) sees engagement as relating to the game play and challenges that arise in playing a game but exist outside the narrative. When compared to immersion, engagement is seen as being the lowest level of involvement on the continuum of engagement, engrossment, and immersion (Brown & Cairns, 2004).

Presence is defined as a "subjective experience of being in one place or environment, even when one is physically situated in another" (Witmer & Singer, 1998, p. 225). Described as such, presence is clearly a concept that could be vastly useful for games research, although it is undoubtedly most applicable to virtual reality experiences. Immersion is an aspect of presence—just one of the forms that presence takes. Similarly, cognitive absorption is described as experiencing a "state of deep involvement with technology" (Zhang, Li, & Sun, 2006, p. 2) and focused immersion is considered to be just a single aspect of cognitive absorption (Agarwal & Karahana, 2000).

As games continue to explode in popularity, with millions of people playing and watching games every day, and the global market reaching $99.6 billion in 2016 (Global Games Market Report, 2016) the need for applied research in this area grows. With the emergence of mixed reality experiences, including virtual reality (VR) and augmented reality (AR) games and experiences, immersion, and presence, or spatial immersion, are increasingly relevant fields of study (Weibel & Wissmath, 2011).

Immersion and appeal have been found to be linearly correlated (Christou, 2014) and immersion has been found to be one of three primary motivations for individuals to engage in online games, alongside achievement and social (Yee, Ducheneaut, & Nelson, 2012).

System Immersion versus Immersive Response

Although there are many varying definitions of immersion, all of which come from different perspectives, there is a clear dichotomy between views of immersion that consider it from the position of the system (*system immersion*), and that of the individual (*immersive response*) (Slater, 1999). The term *immersion,* as a whole, includes all aspects of the experience of being immersed, while *system immersion* is the part of the construct that causes the *immersive response,* the subjective state created into the interaction between the human and the digital element.

Witmer and Singer (1998) view immersion from the standpoint that while investigating immersion, it is important to look not only at the characteristics of an immersive technology system, but also at the individual differences, which may affect the immersive response in any given situation. Their research defines immersion as "a psychological state characterized by perceiving oneself to be enveloped by, included in, and interacting with an environment that provides a continuous stream of stimuli and experiences" (Witmer & Singer, 1998, p. 227). This view depicts immersion as the subjective effect that the use of an immersive system has on an individual user, paying attention to the user's individuality.

Slater (1999) disagrees with Witmer & Singer's (1998) definition of immersion, characterizing the concept instead as "the extent to which the actual system delivers a surrounding environment, one which shuts out sensations from the 'real world'" (p. 1). This depiction of immersion involves the technology of the system in question, and does not focus on the response of the user to the system. Immersive features are "obviously measurable" according to Slater (p. 1). He is of the opinion that it is possible to objectively judge the level of immersion of a technology, the degree to which it provides an immersive response to the participant. Later, Slater (2003) describes immersion as the objective level of fidelity of the sensory stimuli produced by a technological system, again emphasizing the idea that immersion, for him, comes from the technological system.

The immersive response is not directly proportional to the system immersion, as immersion is not a passive response that can be "cued in" by technology. Witmer & Singer's (1998) standpoint is further developed by Norman's (2010) work, which depicts immersion as comprising the immersability of individuals, the tendency of an individual to become immersed in a pursuit, and the immersiveness of games, the ability of a particular game to create an immersive experience. Other factors can influence an individual's immersive response in relation to a given system, and this is discussed in the next section.

Influences on Immersion

In exploring an immersive experience, it should be noted that there are many intrinsic, extrinsic, and mechanical factors that influence the immersive response and these can have a positive or negative effects on the immersive response (summarized in Table 13.1). Factors other than system immersion, external to the system, such as distractions, individual differences, even the time spent, and conscious effort that the user contributes towards an immersive response, can contribute to, or detract from,

Table 13.1 Summary of influences on immersive response as derived from Curran (2013).

Game intrinsic influences	*Game extrinsic influences*
Narrative	Location and mood
Setting	Bridge
Characters and roles	Active occurrences in immersive play
Mechanics	Obscuring the self
	Other individuals
	Familiarity with game

the individual's immersive response. Many studies have examined, and continue to investigate, the factors of a game or system that influence immersion (Curran, 2013; Grimshaw, 2007; Herz, 1997; McMahan, 2003; Norman, 2010).

The narrative, the setting, and the characters and roles presented by the game are all intrinsic elements that influence the creation of the immersive response.

Game narrative is crucial to the experience of immersion, most fundamentally in the game plot and the way that this is expressed, and narrative engagement is defined as the level of players' active engagement in the game (Tychsen, Newman, Brohund, & Hitchens, 2007).

In terms of the setting, consistency is key (McMahan, 2003), although players are often happy to "gloss over" certain minor inconsistencies in order to maintain their sense of immersion (Curran, 2013). Players need not necessarily find a setting inviting in order to become immersed in it: "I need to like the setting." This is not to be confused with "I'd like to live there."

The influence of characters is significant, and depends primarily on how the players relate to the characters. A greater ability to relate to the characters can lead to a greater level of immersion, and ability to relate to characters has been seen to be aided by familiarity with, and similarity to, the characters. As with setting, players need not necessarily like the character in order to immerse but bridging the gap between the player and character stands to greatly enhance immersion. This bridge is facilitated when the player acts, thinks, and, most importantly, feels emotion engendered by the character rather than the player. There are a number of strategies utilized by players in order to become attuned with the character, with parallels to method acting. For many players, empathy with the character was one of the primary motivations for playing the game, and this also enhanced immersion. Emotion did not, in this instance, necessarily imply positive emotion, as immersed players can experience negative emotions during gameplay.

Considering the fact that so many game types involve playing with or against other individuals, it comes as no surprise that the words and actions of other players can have a significant impact on the immersive response, either positively or negatively. In this regard, the caliber of other players, the size of group, and the extent to which trust is established within a group is important for immersion.

A high level of familiarity with the game, its background, and its mechanics, and with the game's characters also facilitates smoothness of play, and therefore, aids the immersive response. The familiarity of all participants in a game allows a consensus to be reached, as decisions are rooted in an agreed reality.

Immersion, when described as being "transported to an elaborately simulated place," is considered to be a pleasurable experience (Murray, 1997, p. 98), and therefore has implications for the level of realism within the "place." It does not follow, however, that total visual and auditory realism in a game is desirable, although this is a widely held misconception concerning immersion in games. There is a common belief that the ability of a virtual environment to deliver an immersive response is related to the ability of the environment to simulate an illusory reality—particularly when it comes to VR games. This belief, termed the *immersive fallacy,* is inaccurate. "According to the immersive fallacy, this reality is so complete that…the player truly believes that he or she is part of an imaginary world" (Salen & Zimmerman, 2004, p. 451). According to Bartle (2003, para. 14), realism can make a virtual environment less compelling: "if you introduce reality into a virtual world, it's no longer a virtual world, it's just an adjunct to the real world. It ceases to be a place, and reverts to being a medium…it takes away that which is different between virtual worlds and the real world: the fact that they are *not* the real world." The level of realism can, at times, detract from an immersive experience rather than add to it (Grimshaw, 2007; Mori, MacDorman, & Kageki, 2012). Bartle concedes that it is possible for immersion to be enhanced by closeness to reality, stating, however, that immersion is "thwarted by isomorphism with it." This reflects a concept introduced by the roboticist Masahiro Mori (Mori et al., 2012), which he termed the *uncanny valley*: the idea that a greater level of realism does not necessarily equate with immersion and that overly lifelike avatars can, in fact, come across as being creepy or scary to a player.

In order to immerse herself in a role, an individual must make an effort, even if she is not consciously setting about doing so. As such, immersion can be a learned technique, as in the case of method acting where actors learn various techniques to enable them to become totally immersed in a role.

Effects of Immersion

As immersion has yet to be definitively described, there is currently little conclusive research on its effects, physical, physiological, or emotional, on the individual player.

During immersive play, it stands to be seen whether or not physiological effects—changes in heartbeat and respiration for example—may be observed during periods of intense activity. Effects of immersion on eye movement have been observed (Jennett et al., 2008) so it stands to reason that there may have other physiological effects associated with a strong immersive response.

It has been found that total immersion can incur positive and / or negative effects, stating that, unlike flow, "immersion is not only viewed as a positive experience: negative emotions and uneasiness (i.e. anxiety) also run high" (Jennett, et al., 2008, p. 658). Some of the positive effects may concern excitement with the game or character, enjoyment on completing tasks, collaboration with other players, whereas negative effects may involve obsession with the game, neglecting real world tasks or relationships, detachment from the real world, and this could hinder a player's progress in real-life situations.

While immersed, a player may feel instant gratification when their character completes a challenge of some type, but conversely, experience negative feelings such as guilt or remorse if their character's action brought about an undesirable outcome.

Other effects can be considered to be *both* positive and negative, depending on the particular situation. Occasionally, when deeply immersed in a game, individuals fail to notice the world around them, becoming unaware of people talking to them, or of the passage of time, and this has been described as becoming "cut off from reality" (Brown & Cairns, 2004, p. 1299). This may provide escapism (Yee, 2007), a healthy diversion from the strains and stresses of everyday life, such as when adolescents "are temporarily transported from life's problems by their playing" (Provenzo, 1991, p. 64). When this behavior is taken to an extreme, however, as in addictive behavior, it can impact negatively on the individual's quality of life.

In more sedentary games, which may involve a more intellectual exercise, a player's health could be jeopardized by overindulging in the game and neglecting physical activity (Rideout et al., 2010). On the other hand, individuals who engage in physical games, for example rhythm games like Dance Dance Revolution or games on the Microsoft Kinect, may have the opposite experience as participation in the game can involve much higher levels of physical activity, exercise, and consequent fitness.

The issue of addiction as an effect of immersion has been considered. Addiction, as a factor in the psychology of video games, has been mentioned as far back as 1983 (Loftus & Loftus, 1983), as a product of the reinforcement effect produced by "beating a previous high score or winning a free game or shooting down enemy spaceships" (p. 14). Game addiction has become a heavily discussed topic in recent years. The degree of immersion while playing a MMORPG (Seah & Cairns, 2008) was discovered to be highly correlated with the addiction score ($r = 0.763$) and addiction was described as "an extreme form of engagement and immersion" (p. 55). This seems to be too simple an explanation, as immersive response is a state of being and an experience of playing, while addiction and engagement are both traits (Seah & Cairns, 2008). A causal link between immersion and addiction to games has yet to be conclusively found.

While there is currently no inclusion of game or online game addiction in the *DSM-IV* (American Psychiatric Association, 2000), or in the *DSM-5* (American Psychiatric Association, 2013), it has been suggested that this is becoming increasingly necessary (Block, 2008; Choi, 2007).

A model that is often used in order to "diagnose" game addiction (Griffiths, 1995) is Brown's addiction model (1997). Interestingly, research (Charlton & Danforth, 2007) suggests that the more peripheral of these criteria (salience, tolerance, and euphoria) occur when an individual is engaged, and occur before the more extreme criteria arise (conflict, behavioral salience, withdrawal symptoms, relapse, and reinstatement). This may indicate that individuals experience a period of high engagement, which is nonpathological, before becoming "addicted," and so addiction should only be diagnosed using the more extreme of Brown's criteria. It may also suggest that engagement or immersion in a game may be a potential starting point for problematic game use, although more research is needed on the topic in order to be conclusive. Although there is no consensus on the addictive personality, it may be considered that some individuals could be more prone to addiction owing to a number of internal or external factors.

The most extreme cases of negative effects of game playing that the media has portrayed were cases of severe bodily deprivation and even death in some cases (Miller & Stanley, 2002). A number of murders (Horowitz, 2007) and suicides have been attributed by the media to games, but the evidence in these cases is inconclusive. The majority of these contentious claims have been discounted by serious commentators, and no evidence has so far been forwarded to substantiate them.

Levels of Immersion

While many individuals share a concept of immersion, it is "not a static experience but describe(s) a *scale* of involvement with a game" (Brown & Cairns, 2004, p. 1300). The development of a taxonomy of levels of involvement is still at an early stage, but what emerges is that when the levels of involvement in a game are considered along a continuum that spans from the lowest to the highest level of involvement, immersion is often regarded as the highest level.

Prominent immersion researchers (Carr et al., 2006; Douglas & Hargadon, 2001) represent immersion and engagement as phenomena that exist along a continuum that depicts the stance of the individual towards the game at any given moment. Immersion and engagement, however, are concepts that are seen, by some, to apply separately to different aspects of computer games (McMahan, 2003). In McMahan's research, immersion itself is believed to apply to the fantasy and narrative aspects of a game, while engagement is seen as relating to the game play and challenges that arise within this play, yet outside the narrative.

Newman, in his work on play sequences (2002, para. 11), depicts two types of engagement as sitting on an "ergodic continuum," with online engagement and offline engagement at opposite ends of the line. Ergodic is defined as "of or relating to a process in which every sequence or sizeable sample is equally representative of the whole" (Newman, 2002). "Online," in this instance, refers to direct input from the player to the game, while "offline" can refer to anything that does not involve player input, such as viewing video cut scenes and score screens at the end of a level. In this regard, Newman sees videogames as being nonergodic, purely because of the existence of game sequences in which a player has no direct input. Calleja (2007), however, disagrees with this view, stating that there are many occasions where players do not engage in a direct input into the system, yet they are stepping back and considering actions, for example, in a strategy game, or waiting in readiness, for example, playing as a sniper in a first-person shooter. There may not be a direct input at every given moment, but the player still has "potential and readiness" (p. 239).

Brown and Cairns (2004) interviewed gamers with the aim of investigating online game experience. Subsequently, they analyzed the data using grounded theory and there emerged a division of immersion into three levels on a continuum—engagement, engrossment, and total immersion. The only games that were found to have an experience of total immersion were first-person perspective games, and games in which the player assumes the role of a character. With each of these levels of immersion, the immersive experience could be constrained by a series of barriers, such as player preferences, game controls, investment, empathy, atmosphere, and game construction. For Brown and Cairns, total immersion is the highest level of immersion on this continuum. This level is described by one of the participants in their grounded theory study as "When you stop thinking about the fact that you're playing a computer game and you're just in a computer" (p. 1299).

In a study that dealt solely with MMORPG development (Alexander, 2005), four levels of immersion were proposed: unimmersed, avatar, character, and persona. Bartle (2010, p. 211) describes the highest level of immersion termed persona in a very clear way: "A persona is a player, in a world. Any separate distinction of character has gone—the player is the character. You're not role playing a being, you are that being; you're not assuming an identity, you are that identity. If you lose a fight, you don't feel that your character has died, you feel that you have died. There's no level of indirection: you are there." This is the level that corresponds with full immersion.

It is possible to divide immersion into mental immersion and physical immersion (Sherman & Craig, 2003). The mental and physical divide of immersion was developed specifically in relation to VR, and as such, this must be kept in mind in interpreting the relevance of these types for general game immersion. Mental immersion is described as being "engaged to the point of suspending disbelief in what they are experiencing" (p. 384) and also as "having 'a sense of presence' within an environment" (p. 9). Physical, or sensory, immersion, is described as "bodily entering into a medium" (p. 9) and is clearly more applicable to VR technology than to online games and other pursuits. Virtual reality research often discusses immersion in these physical terms. Civitarese (2008, p. 281), for example, describes immersion as "the possibility of 'entering' a computer-simulated VR environment and of interacting 'physically,' with the objects inside it, of receiving their responses in real time, thanks to an interface adapted to the characteristics of the human body." Another description of physical immersion, as a "synthetic stimulus of the body's senses via the use of technology" (p. 9) is immediately more easily applicable to the general experience of games.

While other definitions of immersion speak of being "in" the game, for example Bartle's (2005, p. 8) definition of immersion as "the sense that a player has of being in a virtual world" and Murray's (1997, p. 98) description of the concept as "the experience of being transported to an elaborately simulated place," they generally refer to this in the metaphorical sense. Murray follows on to explain that immersion is "a metaphorical term derived from the physical experience of being submerged in water," and the experience of being submerged in an environment is more readily comparable to the emerging VR games space.

Exploring immersion

When researching a theoretical construct such as immersion, it is important that we take into account the "actual perceptions and experiences of the people we plan to study," in this case, gamers (DeVellis, 2003, p. 156). With this in mind, 38 participants, self-identified "gamers," were recruited for a study with the aim of creating a definition of immersion that used the terms of the gamers themselves (Curran, 2013). The majority were male ($n = 32$), with three female participants, and three declining to give a gender specification. Participants were asked to describe and discuss the experiences in which they have felt the most immersed. A qualitative content analysis approach was used to analyze 92 responses, some of which reached over 1000 words, and care was taken to relate emerging concepts to existing literature.

In the definition that emerged, immersion was described as *a subjective state of intense involvement in an imaginary setting, where an individual may either identify with or adopt a character or a setting (or both). If the individual identifies with a character they may adopt some, or all, of the character's attributes. If an individual identifies with a setting some or all of the demands of the setting become the attentional focus. An immersive response can vary in strength and duration, at its most extreme, causing loss of a sense of self or reality and feeling as if one is the character or is "in" the setting.*

It should be noted that the use of the word "setting" here, does not only imply a fully realized imaginary world as is common in real-time strategy games, role-playing games, or first-person shooters, but is also used to apply to more abstract settings such as those in puzzle games like *Tetris*. The use of the descriptor "imaginary," here, does not only imply a fantasy world such as that in *Lord of The Rings*, but can also be used to refer to fictionalized versions of real-world locations, i.e. a game based in New York is still an "imaginary" version of the city, even though it is modeled on a real place.

Types of immersion

During this qualitative exploration of immersion, a category emerged that appeared to indicate a division of immersion into two types of immersive experience, namely, vicarious immersion and visceral immersion (Curran, 2013).

Vicarious immersion has echoes of the idea of make-believe play, which can be seen as emerging in children before their second birthday (Wallace, 1999), and involves being immersed in a role or character, in the thoughts and emotions of that character, and what they experience through the setting and events in the fictional world.

Visceral immersion is the type of immersion associated with feeling a rush of adrenaline, and the feeling of being caught up in the action. Visceral immersion occurs, not in a character or role playing to any extent but when an individual is engaged in the action of a game or film, in the adrenaline rush, an experience of tension, in strategizing and tactics, and yet still managing to be swept away by the experience.

Visceral immersion is often intertwined with the mechanics of a game, involving, among others, falling life points, time running out, and in-game combat. It is possible for a player to become engaged in a game without playing the role of a character—for example, individuals playing the game Super Mario will not picture themselves in the role of a plumber who seeks to save a princess (Clark, 2008), yet they can still become immersed in the experience of playing the game.

These two types of immersion are not, by any means, incompatible or mutually exclusive, and while vicarious immersion is more concerned with role-playing games and visceral immersion is more easily applicable to first-person shooters and action based games, it is possible to experience either type of immersion regardless of game genre and it is also possible that the two types may overlap or occur simultaneously.

Experiencing visceral immersion is possible while in the state of vicarious immersion. This can be seen, for example, in the situation where the character in an online game is caught trying to make his way out of a trap, or during in-game combat while the player is immersed in a character that is desperately fighting for her life. It is debatable whether visceral immersion can occur without some small degree of vicarious immersion, however. It is expected that in puzzle games such as Tetris, for example, that it would be very difficult to experience vicarious immersion: there are no characters present. For this game there is no equivalent action in reality that corresponds with the action during game play, of guiding shapes into spaces, yet visceral immersion occurs readily here as the colored shapes begin to fill up the screen, and time is running out.

In an attempt to offset the lack of generalizability, a subsequent survey study was carried out (N = 217) which clarified the discreteness of vicarious and visceral immersion (Curran, 2013).

Measuring immersion

Immersion research frequently describes it as having a quantitative measurement continuum (Brown & Cairns, 2004; Carr et al., 2006; Douglas & Hargadon, 2001). As such, there are a number of different approaches to measuring elements of the immersive experience.

Lombard et al. (2000) suggest that self-report is the most effective method of measuring the psychological state of immersion.

In a study by Cairns et al. (2006), participants were asked to either carry out an immersive computer game or a nonimmersive, computer-based task. Cairns used savings on *time on task* as a measure for immersion in a game. As part of this research, participants filled in a questionnaire to determine their subjective level of immersion, completed a tangram task, and then were involved in either an immersive or a nonimmersive task, and finally were required to complete a similar tangram task as before. It was found that the experimental group, who were involved in the immersive experience, did not improve their task times as much as the control group. This suggests that immersion may hinder learning or the ability to switch between in-game and real-world situations in rapid succession. Time on task, however, is clearly not an adequate measure of immersion per se, but merely demonstrates the *effect* of trying to carry out a task in rapid succession to being immersed.

The Game Engagement Questionnaire (GEQ—Brockmyer et al., 2009, p. 624) is a "theoretically based measure of engagement in playing video games." In this case, Brockmyer et al. use *engagement* as a "generic indicator of game involvement," explaining that immersion, flow, presence, absorption, and dissociation are more technical terms for the same phenomenon. As such, the GEQ measure is primarily based on existing theory of these concepts.

The Immersive Tendencies Questionnaire (ITQ, Witmer & Singer, 1998) is a tool that is used "to measure the capability or tendency of individuals to be involved or immersed" (p. 230). While Witmer & Singer proposed their instrument to be used for military purposes, Norman (2010) adapted this work and developed two scales, one that measures the immersability of individuals (IAI), and one that measures the immersiveness of games (ING). These also follow the format of a seven–point Likert scale. Both of these instruments are "reliable and…have external validity" (Norman, 2010, p. 7) The ImmersAbility of Individuals Questionnaire, as in the Immersive Tendencies Questionnaire, is a measure of the individual's tendency to become immersed in a pursuit. The ImmersiveNess of Games Questionnaire, in contrast, measures the extent to which a particular game can create an immersive experience or sense of presence.

Measuring the immersive response

Measures exist that indicate the individual's natural *inclination*, *intention or capacity* to immerse herself, as the focus of the ImmersAbility of Individuals questionnaire (IAI, Norman, 2010) or the Immersive Tendencies Questionnaire (Witmer & Singer, 1998). There was no measure found to exist, however, which measured the immersive response during a specified game experience, and so the IMX Questionnaire was developed to fit this purpose and to explore the structure of the immersive response.

The IMX Questionnaire is a quantitative measure of the immersive response in games, developed based on the definition of immersion mentioned above as a subjective state of intense involvement in a game (Curran, 2013). The questionnaire statements were developed from data collected during the initial online discussion (N = 38), and also from qualitative data collected as part of an online survey (N = 217). The IMX Questionnaire takes the form of a summated rating scale with fully anchored five-point Likert-type items and uses agreement as its response type. It was developed using Classical Test Theory (Spector, 1992) and a factor-analysis-based approach. The purpose of factor analysis, a multivariate data reduction technique, is to determine the number of latent variables that underlie the set of items (Bandalos & Finney, 2010; DeVellis, 2003), and this is achieved by investigating these variables that cluster together "in a meaningful way" (Field, 2009, p. 629). The term factor, therefore, refers to a cluster of questionnaire items, with the clusters corresponding to the constructs that "summarize or account for the original set of observed variables" (Hair et al., 2010, p. 92). The aim of factor analysis, for the development of the IMX questionnaire, was to investigate the variables that contribute to the immersive response.

The initial version of the IMX Questionnaire, Version 1 (IMX V1) consisted of 60 items, and Exploratory Factor Analysis (N = 278) suggested the existence of five factors, Vicarious Immersion, Action Visceral Immersion, Mental Visceral Immersion, General Immersion and Group Immersion. The IMX Questionnaire Version 2 (IMX V2) was developed from the results of this Exploratory Factor Analysis, and consisted of 58 items. A Confirmatory Factor Analysis (N = 346) confirmed the existing five factors, although it suggested that a number of items should be removed, and a panel of experts were consulted throughout this process. The IMX Questionnaire Version 3 (IMX V3), which consists of 35 items, has been demonstrated to have high validity and internal consistency reliability (Cronbach's alpha = 0.910) (Nunnally, 1978). The "unending process" of validation is ongoing (Nunnally, 1978, p. 87) but this does not imply that a scale is not ready for use, as long as an acceptable level of validation has been carried out previous to this (Anastasi, 1986, p. 4), and the IMX Questionnaire been used in immersion research (Darzentas et al., 2015).

Overall, 893 participants from 26 countries were involved in the development of this instrument. As is often the case with games research, the majority of participants were male (n ≈ 708), although a significant number of female gamers participated (n ≈ 182).This is a representation of 25% females, although it does not reach the 33% of gamers who are adult females (Entertainment Software Association, 2015), it is more representative than certain game research.

The IMX V3 has three sections. Outlined below is the first section, a set of 16 items, which are applicable to any game. The second set of items is applicable to any game that involves characters, and is made up of 12 items, i.e. items 17–28. The final set of seven items is applicable to multiplayer games, regardless of whether they are cooperative or competitive.

1 I lost myself in the game.
2 I felt caught up in the flow of the game: "in the zone."
3 I kept feeling as if I wanted to play "just ONE more level / mission / hour."
4 I experienced a rush of adrenaline.
5 I was no longer aware of the game "rules," interface or controls, just the game.
6 The game stimulated my reactions (panic, tension, relaxation, suspense, danger, urgency).
7 I got so absorbed that I forgot about time and / or place.
8 During play, I felt at least one of the following: breathlessness, faster breathing, faster heart rate, tingling in my fingers, a fight-or-flight response.
9 I felt as if any action could be my last.
10 I got caught up in tactics and strategy.
11 I got intensely involved in the game.
12 Having to keep up with the speed of the game pulled me in.
13 By succeeding in this game, I felt rewarded and as if I had achieved something.
14 After playing, it took me a moment to recall where I really was.
15 The game was energetic, active and there was a sensation of movement.
16 The game was thought provoking for me.

Five-factor model of immersive response

All analyses indicate that a five-factor model is appropriate for the IMX Questionnaire. The evidence provided by the Cattell Scree plot and the Monte Carlo Parallel Analysis suggested a five-factor model, and this model includes factors that reflect the original definition of immersion and types of immersion, namely, vicarious immersion, action visceral immersion, and mental visceral immersion, as well as including a factor that refers to immersion in group play, group immersion, and a fifth factor, which reflects general immersion. It appears as if visceral immersion, derived as described from the initial qualitative research (Curran, 2013) can be split into two separate types of visceral immersion, which have been termed action visceral immersion and mental visceral immersion.

The general immersion factor describes the most general experience of immersive response: those experiences that occur across game genres and immersive experiences, such as being unaware of the immediate environment outside the game, losing yourself in the game, being "in the zone," feeling drained or disoriented after play, or getting caught up in the flow and wanting to continue playing "just one more level."

Vicarious immersion factor describes the experience of becoming involved in the world of a game, adopting the feelings, thoughts or mannerisms of the character, identifying deeply with the character and feeling as if they are "a real person," and interacting with others from the character's perspective. At its most extreme, the character appears to take on a life of its own: "Sometimes the things I did and said as the character surprised me" and "I became the character." There is a sense of intimacy involved in this type of immersive response.

Action visceral immersion describes the active element of visceral immersion, of being caught up in the action and the excitement of play. It includes items such as "I experienced a rush of adrenaline" and "I felt as if any action could be my last" and an item that describes physical sensations that can accompany this type of immersive response: "During play, I felt at least one of the following: breathlessness, faster breathing, faster heart rate, tingling in my fingers, a fight-or flight response."

Mental visceral immersion includes items that describe excitement, yet it leans towards the aspect of visceral immersion that describes becoming "engaged in strategizing and tactics," finding the game thought provoking and being intensely involved in the game. Mental visceral immersion also has an element that references feeling rewarded and a sense of achievement from succeeding in a game goal, as well as a sense of control and impact over what occurs.

Group immersion describes the aspects of the immersive response, which can only occur when playing with other individuals, as is common in competitive or cooperative online games, such as camaraderie and feeling a sense of involvement that matched the involvement of other players.

Conclusion

When exploring immersion, it is essential to take into account that there are differing definitions of immersion depending on discipline, technology, and perspective and that there is no one single widely accepted definition of immersion. Immersion is split into immersive response and system immersion, and the effects of immersion and influences on immersion continue to be explored. Immersion is considered to comprise a continuum of levels of involvement, from low to high involvement, and as such it is possible to quantitatively measure or detect immersion in a game session, as well as the immersive tendencies and factors of a game that contribute to an immersive response. Beyond this, it should be considered when researching immersion that there are different types of immersive experience that are depicted in the proposed five-factor model of immersion: general immersion, vicarious immersion, action visceral immersion, mental visceral immersion and group immersion.

Overall, much of the discussion in this chapter has focused on immersion in games, as the area where the majority of research lies, but the relevance of immersion in other pursuits, such as working, studying, consuming media, and others cannot be overstated, particularly when these pursuits are looked at through the lens of human-computer interaction. The experience of paying bills may not be particularly engaging, but with the application of the right interface or app, it is possible to make even the most mundane tasks more immersive.

References

Agarwal, R., & Karahana, E. (2000). Time flies when you're having fun: Cognitive absorption and beliefs about information technology usage. *MIS Quarterly, 24*(4), 665–694.

Alexander, T. (2005). *Massively Multiplayer Game Development 2*. Rockland, MA: Charles River Media Inc.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders DSM-IV-TR* (4th ed.). Washington, DC: Author.

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders DSM-5* (5th ed.). Washington, DC: Author.

Anastasi, A. (1986). Evolving concepts of test validation. *Annual Review of Psychology, 37*(1), 1–16.

Bandalos, D. L., & Finney, S. J. (2010). Factor analysis: Exploratory and confirmatory. In G. R. Hancock & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 93–114). New York, NY: Routledge.

Bartle, R. (2003, July 28). Not yet you fools! *Game+Girl=Advance*. Retrieved from http://www.gamegirladvance.com/2003/07/not-yet-you-fools.html

Bartle, R. (2005). Virtual worlds: Why people play. In: T. Alexander (Ed.), *Massively multiplayer game development 2*. Rockland, MA: Charles River Media Inc.

Bartle, R. (2010). Players. In H. Donelan, K. Kear & M. Ramage (Eds.), *Online communication and collaboration: A reader* (pp. 210–212). New York: NY: Routledge.

Block, J. J. (2008). Issues for *DSM-V:* Internet addiction. *American Journal of Psychiatry 165*(3), 306–307.

Brockmyer, J. H., Fox, C. M., Curtiss, K. A., McBroom, E., Burkhard, K. M., & Pidruzny, J. N. (2009). The development of the Game Engagement Questionnaire: A measure of engagement in video game-playing. *Journal of Experimental Social Psychology, 45*(4), 624–634.

Brown, E., & Cairns, P. (2004). A grounded investigation of immersion in games. In *Proceedings of the ACM Conference on Human Factors in Computing Systems, Vienna, Austria: CHI 2004* (pp. 1297–1300). New York, NY: ACM.

Brown, I. (1997). A theoretical model of the behavioural addictions applied to offending. In J. E. Hodge, M. McMurran & C. R. Hollin (Eds.), *Addicted to crime?* (pp. 15–63). Chichester: John Wiley & Sons, Ltd.

Cairns, P., Cox, A., Berthouze, N., Dhoparee, S., & Jennett, C. (2006, July). Quantifying the experience of immersion in games. In *Cognitive Science of Games and Gameplay Workshop, conducted at the 28th Annual Conference of the Cognitive Science Society*. Red Hook, NY: Curran Associates Inc.

Calleja, G. (2007). Digital game involvement: A conceptual model. *Games and Culture, 2*(3), 236–260.

Carr, D., Buckingham, D., Burn, A., & Schott, G. (2006). *Computer games: Text, narrative and play*. Cambridge: Polity Press.

Chapman, P. (1997). *Models of engagement: Intrinsically motivated interaction with multimedia learning software* (Unpublished master's thesis). University of Waterloo, Waterloo, ON, Canada.

Charlton, J. P., & Danforth, I. D. W. (2007). Distinguishing addiction and high engagement in the context of online game playing. *Computers in Human Behavior, 23*(3), 1531–1548.

Choi, Y. H. (2007). *Advancement of IT and seriousness of youth Internet addiction*. In National Youth Commission, 2007 International Symposium on the Counseling and Treatment of Youth Internet Addiction. Seoul, Korea.

Christou, G. (2014). The interplay between immersion and appeal in video games. *Computers in Human Behavior. 32*, 92–100.

Civitarese, G. (2008). Immersion versus interactivity and analytic field. *International Journal of Psychoanalysis, 89*(2), 279–298.

Clark, B. R. (2008). *Using the ZMET method to understand individual meanings created by video game players through the player-Super Mario avatar relationship*. (Unpublished master's coursework). Brigham Young University, Provo, UT.

Coomans, M. K. D., & Timmermans, H. J. P. (1997). Towards a taxonomy of virtual reality user interfaces. In *Proceedings of the International conference on Information Visualisation (IV97)* (pp. 279–284). Los Alamitos, CA: IEEE Computer Society.

Csíkszentmihályi, M. (1990). *Flow: The psychology of optimal experience*. New York, NY: Harper & Row.
Curran, N. (2011). Stereotypes and individual differences in role-playing games. *International Journal of Role-Playing, 2*, 44–58.
Curran, N. (2013) *The psychology of immersion and development of a quantitative measure of immersive response in games* (Unpublished doctoral thesis). University College Cork, Cork, Ireland.
Darzentas, D., Brown, M., & Curran, N. (2015). Designed to thrill: Exploring the effects of multimodal feedback on virtual world immersion. In Kurosu, M. (Ed.), *Human-computer interaction: Users and contexts.* (pp. 384–395). Los Angeles, CA: Springer.
DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Thousand Oaks, CA: Sage.
Douglas J. Y., & Hargadon, A. B. (2001). The pleasures of immersion and engagement: Schemas, scripts and the fifth business. *Digital Creativity, 12*(3), 153–166.
Entertainment Software Association. (2015). *2015 essential facts about the computer and video game industry*. Retrieved from http://www.theesa.com/wp-content/uploads/2015/04/ESA-Essential-Facts-2015.pdf
Field, A. P. (2009). *Discovering statistics using SPSS* (3rd ed.). London: ISM, Sage.
Global games market report. (2016, April). *Newzoo*. Retrieved from https://newzoo.com/insights/articles/global-games-market-reaches-99-6-billion-2016-mobile-generating-37/
Griffiths, M. (1995). Technological addictions. *Clinical Psychology Forum, 76*, 14–19.
Grimshaw, M. (2007). Sound and immersion in the first-person shooter. *International Journal of Intelligent Games and Simulation, 5*(1), 119–124.
Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis: A global perspective* (7th ed.). London: Pearson Education.
Haywood, N., & Cairns, P. (2005). Engagement with an interactive museum exhibit. *People and Computers, 19*, 113.
Herz, J. C. (1997). *Joystick nation: How video games ate our quarters, won our hearts and rewired our minds*. Boston, MA: Little, Brown, & Co.
Horowitz, S. J. (2007). Competing Lockean Claims to Virtual Property. *Harvard Journal of Law and Technology, 20*(2), 443–458.
Jennett, C., Cox, A., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. (2008). Measuring and Defining the experience of immersion in games. *International Journal of Human-Computer Studies, 66*(9), 641–661.
Laurel, B. (1993). *Computers as theatre* (2nd ed.). Reading, MA: Addison-Wesley.
Loftus, G. R., & Loftus, E. F. (1983). *Mind at play: The psychology of video games*. New York, NY: Basic Books.
Lombard, M., Ditton, T. B., Crane, D., Davis, B., Gil-Egui, G., Horvath, K., & Rossman, J. (2000). *Measuring presence: A literature-based approach to the development of a standardized pencil-and-paper instrument*. Abstract submitted for presentation at presence 2000: the Third International Workshop on Presence.
McCarthy, J., & Wright, P. (2004). *Technology as Experience*. Cambridge, MA: MIT Press.
McMahan, A. (2003). Immersion, engagement and presence: A method for analysing 3D video games. In M. J. P. Wolf & B. Perron (Eds.), *The video game theory reader* (pp. 67–86). New York, NY: Routledge.
Miller, S. A., & Stanley, A. (2002, March 31). Death of a game addict. *Milwaukee Journal Sentinel*.
Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley. *Robotics and Automation Magazine, 19*, 98–100.
Murray, J. (1997). *Hamlet on the holodeck*. Cambridge, MA: MIT Press.

Newman, J. (2002). The Myth of the ergodic videogame. *Gamestudies: The International Journal of Computer Game Research, 2*(1), 1–17. Retrieved from http://www.gamestudies.org/0102/newman/

Norman, K. L. (2010). *Development of instruments to measure immersability of individuals and immersiveness of video games.* Retrieved from HCIL website: http://hcil2.cs.umd.edu/trs/2010-33/2010-33.pdf

Nunnally, J. C. (1978). *Psychometric theory* (2nd ed.). New York, NY: McGraw Hill

Pine, B. J., & Gilmore, J. H. (1999). *The experience economy: Work is theater and every business a stage.* Boston, MA: Harvard Business School Press.

Provenzo, E. (1991). *Video kids: Making sense of Nintendo.* Cambridge, MA: Harvard University Press.

Rawlinson, G. (1861). *The history of Herodotus.* A New English Version, New York: D. Appleton.

Rideout, V., Foehr, U. G., & Roberts, D. (2010). *Generation M^2: Media in the lives of 8- to 18-year olds.* Menlo Park, CA: Henry J. Kaiser Family Foundation.

Salen, K., & Zimmerman, E. (2004). *Rules of play: Game design fundamentals.* Cambridge MA: MIT Press.

Schubert, T., & Crusius, J. (2002). Five theses on the book problem: Presence in books, film and VR. In *Presence 2002: Proceedings of the Fifth International Workshop on Presence* (pp. 53–59). Porto, Universidade Fernando Pessoa.

Seah, M., & Cairns, P. (2008). From immersion to addiction in videogames. In *Proceedings of the 22nd British HCI Group Annual Conference on HCI 2008: People and Computers XXII: Culture, Creativity, Interaction* (vol. 1, pp. 55–63). Liverpool: British Computer Society.

Sherman, W. R., & Craig, A. B. (2003). *Understanding Virtual Reality: Interface, Application and Design.* San Francisco, CA: Morgan Kaufmann Publishers.

Slater, M. (1999). Measuring presence: A response to the Witmer and Singer presence questionnaire. *Presence: Teleoperators and Virtual Environments, 8*(5), 560–565.

Slater, M. (2003). A note on presence terminology. *Presence-Connect, 3.* Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.800.3452&rep=rep1&type=pdf

Spector, P. (1992). *Summated Rating Scale construction: An introduction.* Newbury Park, CA: Sage Publications.

Tychsen, A., Newman, K., Brohund, T., & Hitchens, M. (2007). Cross-format analysis of the gaming experience in multi-player role-playing games. In *Proceedings of the 2007 DiGRA International Conference, Tokyo, Japan: Situated Play* (vol. *4*, pp. 49–58). Tokyo, Japan: DiGRA, Japan.

Wallace, P. (1999). *The psychology of the Internet.* New York, NY: Cambridge University Press.

Weibel, D., & Wissmath, B. (2011). Immersion in computer games: The role of spatial presence and flow. *International Journal of Computer Games Technology, 2011*, 6.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments: A presence questionnaire. *Presence, 7*(3), 225–240.

Yee, N. (2007). Motivations of play in online games. *Journal of Cyber Psychology and Behaviour, 9*(6) 772–775.

Yee, N., Ducheneaut, N., & Nelson, L. (2012). Online Gaming Motivations Scale: Development and validation. In *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI 2012)* (pp. 2803–2806). New York, NY: ACM.

Zhang, P., Li, N., & Sun, H. (2006). Affective quality and cognitive absorption: Extending technology acceptance research. In *Proceedings of the 39th Annual Hawaii International Conference on System Sciences* (vol. 8, p. 207). Piscataway, NJ: IEEE.

Part IV
Evaluation Methods

14

Usability Testing

Sirpa Riihiaho

Introduction

Usability is a multifaceted term. In user-centered product development, it normally means the quality of use in a context, or "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use", as the ISO 9241-11 standard (1998) defines it. In user-centered design, it is essential to evaluate usability throughout the whole development process instead of massive tests only at the end.

Usability evaluations can be divided into *formative and summative* evaluations according to the goals of the evaluation. Formative evaluation is used to gather user feedback for further development, whereas summative evaluation is used to assess if the set usability requirements are met (Hewett, 1986). As usability should be assessed several times in the development process, most evaluations are formative by nature, allowing more room for modifications in the methods and the test settings.

The usability evaluation methods can also be divided according to the user involvement into user testing methods and usability inspection methods. The term usability testing is sometimes used as a synonym for user testing or usability evaluation. To make a distinction between a method and a group of methods, this chapter uses the terms as follows:

- *User testing* covers a group of usability evaluation methods that involve user participation, and
- *Usability testing* is a user testing method in which one or more representative users at a time perform tasks or describe their intentions under observation.

The process of usability evaluation with usability testing has four phases:

1. Design and preparation of the tests.
2. Conducting the test sessions.
3. Analyzing the results.
4. Communicating the results.

The Wiley Handbook of Human Computer Interaction, Volume 1, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

Although there are several variations of usability testing, tests with predefined test tasks in controlled test environments are still a general practice. However, tests in the field or on the move are becoming increasingly common. This chapter gives an overview of the basic process of usability testing, and then presents some modifications of the traditional usability test.

Planning Usability Tests

A test plan is used to obtain approval from the management and other people involved, and to clarify the goals of the test. When talking about goals for usability evaluation, there are two different levels of goals: the overall goal and motivation for the evaluation process, and the more specific goals for the selected usability attributes. Motivations for usability evaluation include ensuring that the product reaches its minimum level of usability; getting feedback of how well the objectives are met; and identifying potential usability defects in the product. Testing with users also gives an opportunity to collect new information about the users' needs, to compare the design with competing products, and to find issues for training.

Within these general goals and motivations, there are usually some more specific goals for evaluation. These goals should be specific to ensure that the evaluations truly assess the attributes of interest, and that a suitable set of evaluation methods, test participants, and test tasks are selected. For example, in a commercial off-the-shelf (COTS) software system procurement in which our university was purchasing a new Current Research Information System (CRIS), our general goal was to find the best system out of the two alternatives, emphasizing its costs, utility, and usability (Riihiaho, Nieminen, Westman, Addams-Moring, & Katainen, 2015). Several universities were involved in the procurement process that our university organized, so it was a very attractive deal for the competing companies. The winning system would be purchased by several Finnish universities, so it was a major investment, which could lead to litigation from companies faring badly in the comparison. We considered quantitative results easier to justify in court than qualitative ones, so we emphasized statistical results and comparisons. This requirement affected many decisions we made during the process of testing. We will now consider the questions asked when designing the tests.

Must we have something functional to be tested?

It is easier for the test users to comment on the system if they can try it out. This is especially important in summative testing, so a functional prototype or a working system is preferred if statistical analysis is required. In formative evaluation, however, a more conversational interaction between the test user and the test moderator is appropriate, so the moderator can help the test user to use a simple sketched prototype or a website wireframe.

When using paper prototypes, it is recommended that more than one version is used, as the study by Tohidi, Buxton, Baecker, and Sellen (2006a) shows that users give significantly higher ratings when they assess only one design compared to assessing three alternatives. The number of redesign proposals, however, remained quite low in both settings. To get more ideas, Tohidi, Buxton, Baecker, and Sellen (2006b) asked the test participants to sketch their ideal product design on a sheet of paper after testing the prototypes in the previous study. If the test users had seen multiple versions of the system, their own sketches were more versatile.

If a paper prototype and a running system are compared, for example when comparing new ideas to the old system, one must remember that subjective ratings are not quite comparable. Sauer and Sonderegger (2009) manipulated both the aesthetics and the fidelity of the prototypes. Somewhat surprisingly, the fully functional system with lower aesthetics received significantly lower ratings for attractiveness compared to all the other versions, including the similar paper prototype. The researchers gathered that the paper prototypes left room for the users to figure out the final looks of the system, so the test users may have made their assessments based on these implications instead of the prototypes in hand. (Sauer & Sonderegger, 2009).

Which metrics should we use?

The goals of the test should be specified right from the beginning to be able to focus the tests and to limit the scope to certain user groups, attributes, and tasks. For each attribute of interest, there should be measures set to be able to evaluate the level of usability. Hornbæk (2006), for example, gives an excellent summary of different metrics for effectiveness, efficiency, and satisfaction, such as binary task completion, accuracy, error rates, completeness, and quality of outcome.

These usability metrics are important, especially in summative evaluation, but can also be useful in formative evaluation to give quantitative results for comparisons with former tests and competing products. However, if the thinking-aloud method is used in the tests, the measurements should emphasize metrics other than performance times, as thinking aloud may slow down the performance.

In our CRIS case, we selected effectiveness, efficiency, and satisfaction from the ISO 9241-11 standard (1998) as our measurable attributes. Effectiveness was measured by how many tasks were completed successfully within the maximum time of 15 minutes. If the task was not finished in time, was abandoned by the user, or was determined as not successful either by the user or the evaluators, it was not counted. Efficiency was measured as the average time for a user to complete the tasks, and the maximum 15 minutes was used for unsuccessful tasks. Since performance times were measured, we did not use thinking aloud in these tests. Finally, satisfaction was measured with the positive version of the System Usability Scale (SUS) (discussed later on in this chapter), and the users were asked to pick their favorite system after testing both the systems (Riihiaho *et al,*. 2015).

Whatever the metrics are, they should include both objective and subjective measures but still be clearly separated. For example, the time to complete a task does not generally vary a lot between the test users, but the way the users perceive the required time may vary much more, and may reveal frustrations in the use of the system (Hornbæk, 2006).

What sort of, and how many, test users should we have?

The participants in a usability test should represent the real users as well as possible. Friends, family or coworkers are not preferred, because a close relationship between a participant and a moderator can easily bias the results. Schrier (1992) also points out that it is usually easier for a test user to interact with a moderator and observers, and to criticize the product, if the observers are not acquainted. However, in East Asian cultures, it is recommended that the test users are familiar with the moderator, and that they are of higher status than the moderator to feel comfortable in giving negative comments on the evaluated system (Yeo, 2000).

There are many factors that should be considered when selecting a suitable set of test participants, such as various user groups, age, gender, and level of expertise. The users' expertise is affected by their experience with the evaluated system or its previous versions; their experience with computers, information technology, and technical devices in general; and their experience with the task domain (Nielsen, 1993, p. 43). The users' organizational knowledge, training, input device skills, qualifications, and language skills also affect their previous experience (Maguire, 2001).

Especially in evaluating occupational systems, the test users should be experts in the task domain, but the suitable level of experience with similar systems should be considered according to the goals of the test. Novice users are good at revealing problems with learnability and affordance, whereas expert users are good at revealing inconsistencies with similar products, and finding illogical features. In our CRIS case, we decided to use both expert users from our administration, and novice users from our researchers and postgraduate students (Riihiaho *et al.,* 2015).

If the system to be evaluated is intended for children, it is a good idea to do the testing in two phases: in the first phase the test team introduces itself to the children, and only in the second phase, the children are asked to perform some tasks under observation. With children, the test participants' capacity and tendency to verbalize, ability to concentrate, motivation, ability to adjust to strange environments, trustworthiness of self-report, as well as their knowledge and skills must be considered (Markopoulos & Bekker, 2003). In school environment, the teachers are valuable assistants in choosing suitable test participants.

The required number of test users has been a popular topic, although there are several studies showing no relation between the number of test users and the number of detected usability problems, such as the study by Molich, Ede, Kaasgaard, and Karyukin (2004). Furthermore, the analysis by Lindgaard and Chattratichart (2007) shows that the number of test tasks correlates with the number of detected problems better than the number of test users.

Even so, the desired coverage of the evaluation, complexity of the system, planned number of iterations, available test resources, and the need for statistical analyses affect the required number of test users. In formative testing, five users are often used, as the early studies by Lewis (1994), Nielsen (1994a), and Virzi (1992) showed that about 80% of the usability problems could be found with five users—possibly due to assessing rather simple systems. In summative testing and statistical results, more users are often needed. To estimate the required number of users, Lewis (2001) gives a basic formula for estimating the proportion of problems discovered: $1 - (1 - p)^n$, where p is the problem-discovery rate, and n is the number of test users. In a more recent meta-analysis by Hwang and Salvendy (2010), nine users were usually required for the 80% coverage of the problems.

In our CRIS case, we were not interested in the proportion of the problems found, but hoped for some statistical differences in the test results. We therefore used the results of six expert and six novice users in the comparisons. To get these

12 comparable results, we needed three extra users, as we had some technical problems during these three tests that biased their results. In addition, we had one pilot test user from both groups to check our test settings before the actual tests.

What test tasks and scenarios should we select?

When designing the usability tests, it is recommended to make one or two usability inspections to get some early notions on the usability problems. If the inspections reveal some considerable problems, they can be fixed before user involvement, or validated with the test users with some test tasks.

Interviewing a few users in real use contexts is a rewarding method when planning usability tests. Firstly, it gives ideas for suitable scenarios and test tasks. Secondly, it gives a deeper understanding of the users' goals, and thereby helps to find a focus for the tests. Finally, information about the users' goals and priorities helps to produce effective redesign proposals to solve the problems uncovered in the tests.

The test tasks should be realistic and represent the actual use expected in the field. The tasks should cover the most important parts of the product, and reflect the attributes selected as the focus of the test. The tasks should not last more than an hour to keep the test users focused. However, the coverage of the tasks should be wide enough, and the variety of the tasks big enough to give the users a chance to try the system out. Moreover, the analysis by Lindgaard and Chattratichart (2007) showed that the number of problems detected correlates with the number of test tasks, so it is recommended to do several tasks with each user, as the effort of getting test users is often considerable.

The quality and relevance of the test tasks affect the number of problems found (Skov & Stage, 2012). Thus, the test tasks should be meaningful, having a clear and unambiguous goal, but with no instructions on how to complete them. Even the wording of the test tasks should not give distinct hints of the required actions.

Having an easy task at the beginning of the test gives the test users a chance to become familiar with the test settings and to relax somewhat. The tasks should be independent from each other, and presented one at a time, so that some tasks could be skipped unnoticed if time is running out. Scenarios can be used to make a connection between the tasks, and to give the participants a social and organizational context for the tasks, such as a specific shift in nurses' work. People from a Chinese culture may find isolated tasks artificial and hard to understand without reasonable background scenarios (Clemmensen *et al.*, 2009).

In our CRIS case, we had five test tasks for both user groups. For the researchers and postgraduate students, the tasks were the following (Riihiaho *et al.*, 2015):

- enter the article details manually in the system;
- create a description of a research visit to another university;
- modify your research profile;
- construct a CV with publications and activities from your profile;
- import previous publications to the system from a BibTeX file.

The maximum length for each task was 15 min, after which the users were asked to move on to the next task. The test users could use their own research information in the tasks, as we had entered their data into the compared systems before the tests. With this realistic data, the users had an opportunity to assess the utility of the system in addition to its usability, as Bødker and Madsen (1998) suggest.

What should we ask in questionnaires and interviews?

Questionnaires are often used both before and after the users perform the test tasks. Usually, the pretest questionnaires focus on the users' background information as well as their expectations and attitudes toward the system. Posttest questionnaires, then, gather information about the users' experiences with the evaluated system. Structured and semistructured interviews can also be used to supplement or replace these posttest questionnaires.

It is recommended that standardized questionnaires are used instead of designing one's own questionnaires, due to their better reliability (Hornbæk & Law, 2007), especially in summative evaluation. The System Usability Scale (SUS, https://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html), for example, was made freely available in 1986, and has subsequently become almost a *de facto* standard in usability evaluation (Brooke, 2013). The SUS questionnaire consists of ten statements that are rated on a five-point Likert-type scale (strongly disagree versus strongly agree). The combined results are presented as a single number "representing a composite measure of the overall usability of the system being studied" (Brooke, 1996). This score ranges from 0 to 100, making 50 points the theoretical mean score, but the actual realized average score is close to 70 (Bangor, Kortum & Miller, 2008).

The SUS questionnaire includes five positive and five negative statements (Brooke, 1996). A mixture of positive and negative statements is used to balance various biases, and to make respondents think about each statement before answering. However, negative statements make mistakes more likely both in answering the questions and in interpreting the results, so Sauro and Lewis (2011) have developed an all-positive version of the SUS questionnaire. Since their results are almost identical with the original questionnaire, it is recommended that this positive version is used, especially if there is no possibility for the evaluators to check the answers with the users. To avoid misinterpretations, we also used the all-positive version in our CRIS case with a minor modification: Sauro and Lewis (2011) used the term website in their version, so we returned the term system in our version. The original and modified versions of the SUS statements are presented in Table 14.1.

Another commonly used standardized and freely available questionnaire is the NASA Task Load Index (NASA-TLX, http://humansystems.arc.nasa.gov/groups/tlx/). It is designed to assess the subjective workload in human-machine interaction using six subscales: mental, physical, and temporal demands, as well as own performance, effort, and frustration. Commercial questionnaires include the Software Usability Measurement Inventory (SUMI, sumi.uxp.ie), and the Website Analysis and MeasureMent Inventory (WAMMI, http://www.wammi.com/index.html).

When using questionnaires, one should remember that people are eager to please in surveys and interviews. Potentially sensitive questions should therefore be asked through the computer rather than in person to minimize the bias, and the value of 3.6 should be used as an estimate for a neutral mean instead of the mathematical 3.0 when using a scale from 1 to 5 (Nielsen, 1993, pp. 37, 213–214). Furthermore, the test users' subjective quality judgments in posttest questionnaires do not necessarily reflect

Table 14.1 Original and modified positive version of the SUS statements.

	Original SUS statements	*Modified all-positive statements*
1	I think that I would like to use this system frequently	I think that I would like to use this system frequently
2	I found the system unnecessarily complex	I found the system to be simple
3	I thought the system was easy to use	I thought the system was easy to use
4	I think that I would need the support of a technical person to be able to use this system	I think that I could use this system without the support of a technical person
5	I found the various functions in this system were well integrated	I found the various functions in this system were well integrated
6	I thought there was too much inconsistency in this system	I thought there was a lot of consistency in this system
7	I would imagine that most people would learn to use this system very quickly	I would imagine that most people would learn to use this system very quickly
8	I found the system very cumbersome to use	I found the system very intuitive
9	I felt very confident using the system	I felt very confident using the system
10	I needed to learn a lot of things before I could get going with this system	I could use this system without having to learn anything new

the whole test, but only its most recent incidents (Hassenzahl & Sandweg, 2004). Therefore, it is recommended that specific posttask questionnaires should be used if task-relevant information is required (Sauro & Lewis, 2009).

Children are especially challenging participants in interviews and questionnaires, because their language and motor skills, reading age, confidence, self-belief, and desire to please affect their answers (Read & MacFarlane, 2006). When we were evaluating an educational system for children, one method we used was to let five children test the functional prototype as a group in a classroom setting with their own teacher. After this testing, we interviewed the children also in a group with the help of a feedback game that we developed.

The feedback game used a physical gameboard having eight rows with interview questions and five columns with smiley faces, and physical tokens that the children could set to this board after each question. For example, we asked if the children found the system fun to use. All the children set one of their tokens to the smiley columns at the same time, and then the moderator started a discussion by asking a more detailed question to clarify the reasons for the answers. The question about fun, for example, was coupled with a question about what was the most fun or least fun thing about the system. The physical objects and the rules saying that everyone had their own turn to explain the answer made the interaction easy and quite natural for the children (Kantosalo & Riihiaho, 2014).

Where should we test?

Usability tests can be conducted practically anywhere. The customer's site is more familiar to the participants, making it easier for them to relax, but it is more challenging for the evaluators, because interruptions are hard to control, and the equipment has to be brought along. Specific laboratories, then, give greater control of the variables of interest, and the measurements are more precise, but the artificial environment can produce unrealistic results. For example, McDonald, Monahan, and Cockton (2006) estimate that as many as two-thirds of the problems identified in their contextual interviews are related to the context of use.

Despite the physical location of the test, the test context should include the most critical components of the real use context. Issues to be considered include realistic test data; having everyday materials and tools available; the need for simulated interruptions and time pressures; the need for cooperation between users; the placement of the system and other relevant material; as well as the organizational situation and the role of the user (Bødker & Madsen, 1998). For our CRIS case, our usability laboratory provided a well-controlled environment, close to the users, and real user-specific research data was used to bring more reality into the use context for each test user.

Although the use context has numerous attributes in several levels of detail, the most common contexts easily cover over half of the use situations. For example, Kim, Kim, Lee, Chae, and Choi (2002) studied the use contexts of mobile Internet. They categorized the use contexts with eight parameters: reason to use the system; emotional status; one or two hands used; moving or staying still; visual distractions; auditory distractions; number of people around; and the level of interaction with others. Their results showed that just two types of use contexts out of the 256 theoretical possibilities covered over 20% of the reported use sessions. In addition, the availability of hands, movement of legs, and the number of people around the user had the biggest impact in the detected usability problems in the mobile Internet. (Kim *et al.*, 2002).

Should we have a moderator present or not?

Having a moderator in the test room has both advantages and challenges: the integrity of the test data is less biased if the moderator is not present, but the user may feel uncomfortable alone in the test room, and this in turn, may bias the results (Dumas & Loring, 2008, pp. 125–131). Guidelines for interacting with the test users emphasize the risk of biasing the users' behavior by the moderator's tone of voice and body language, but there are quite little studies of the actual effects of the moderator presence, and with diverse results.

Negative effects have been quite rare, as the studies by Held and Biers (1992) are the only ones revealing more negative system ratings from expert users if a moderator was present, but even in these studies, novice users gave slightly positive ratings. Other studies, then, have indicated more positive effects, such as encouraging participants to search for more information than on their own, and completing the test with more certainty (Schulte-Mecklenbeck & Huber, 2003). The study by Sonderegger and Sauer (2009) also suggested that a good rapport between the moderator and the test user could enhance the users' performance. Finally, the author's own experiment (Riihiaho, 2015) indicated that the presence of a test moderator caused significantly more positive results when the users assessed how pleasant the evaluated system was in a posttest questionnaire.

As the moderator's presence may affect the users' performance and subjective ratings, it is recommended that the users are left alone in summative testing, or at least to minimize interaction with the users. For example, in our CRIS comparison, we had a moderator present in the tests to help with technical problems, but the moderator remained mostly silent, and gave only predefined replies when needed. In formative testing, though, it is highly recommended to have a moderator next to the users, as this gives a valuable opportunity to observe the users' first impressions closely, and to probe their expectations and experiences when the events are fresh in mind (Riihiaho, 2015).

Should we use thinking aloud or not?

In a usability test, the test users are usually asked to think aloud while performing the test tasks. Thinking aloud has been used in psychological studies for a long time as a method to study cognitive processes, and it has become one of the central methods in usability testing, too.

The participants may verbalize their thoughts concurrently when performing the tasks, or describe their thoughts retrospectively after completing the test tasks. Retrospective reports are viewed as less useful and less reliable than concurrent reports, as they rely on participants' memories of what they have been thinking some time ago (Ericsson & Simon, 1980). Retrospective reports also require more time from the test users, so concurrent thinking aloud has become the most general form of verbal reports in usability testing.

There are three levels of verbal reports: at level 1, the information is told in the same way as it is processed in the working memory; at level 2, the original information is not in verbal form, such as an image, so it has to be translated; and at level 3, the subjects are asked to do something more than just telling their thoughts aloud, such as selecting information according to given instructions. If the subjects are asked to describe their motor activities or routine actions to which they would not otherwise pay attention, the verbalization falls into level 3. At levels 1 and 2, the cognitive processes remain the same as if the participants acted silently, but level 2 may still slow down the performance. However, at level 3, the subjects may alter their normal behavior, and pay more attention to information, which can make them more efficient in the present or following tasks (Ericsson & Simon, 1980).

In summative evaluation, performance times are often used to measure the level of usability. Thinking aloud should therefore not be used in summative evaluation, or it needs to be kept at levels 1 or 2. In formative evaluation, however, it is important to get as much information from the users as possible, so it is common to ask the users to give reasons for their actions instead of just stating them. The evaluator may also probe for further information between the tasks and even during a task if something surprising occurs, or if the user gets stuck while performing the task. This relaxed thinking aloud gives room for discussions and also helps create a good rapport in the test session. In our CRIS case, we did not use thinking aloud as we used performance times as our central usability metrics.

Conducting Test Sessions

A very practical list of ten steps for conducting the test sessions (Gomoll, 1990) includes:

1. Introduce yourself.
2. Describe the purpose of the test in general terms.

3 Tell the participants that they may quit at any time, and still get a fee if such is available.
4 Explain the purpose of the equipment in the room.
5 Explain how to think aloud, and give an example.
6 Explain that you cannot provide help during the test.
7 Describe the tasks and introduce the product.
8 Ask if the user has any questions, and then begin the observation.
9 Conclude the observation.
10 Use the results.

Before any actual tests, at least one pilot test is needed to check the test tasks, instructions, equipment, and placements. The pilot test should be made early enough, so that the evaluators still have enough time to make changes if needed. The pilot user does not have to be from the target group but someone outside the test team to be able to spot ambiguous wordings or illogical task sequences. As we had two user groups with different test tasks in our CRIS case, we ran one pilot test with both groups a few days before the actual tests.

Usability tests are typically conducted to make products less stressful to use, but the testing process itself can be very stressful for the test participants (Schrier, 1992). The main ethical considerations before the test session include ensuring that everything is ready before the participant arrives; informing the participant about the state of the system and of the confidentiality of the results; and ensuring that the participant knows that it is the product that is tested, not the user (Nielsen, 1993, p. 184).

To show that the moderator honors the users' skills and knowledge, the roles of a master and an apprentice are appropriate for the test user and the moderator, similar to the roles in contextual inquiry by Beyer and Holtzblatt (1995). In no circumstances may the moderator indicate that the participant is making mistakes or proceeding too slowly. Instead, the moderator should "maintain a positive attitude throughout the entire test session, no matter what happens" (Schrier, 1992). After the test, the moderator should thank the participants for their help, and ensure that they stay anonymous in the results. The recordings are presented outside the testing team only with the participants' permission.

Analyzing and Communicating Results

The test analysis gives interpretations of what happened in the test sessions, and what problems and successes emerged. The problems should be organized by their importance—scope and severity (Dumas & Redish, 1993, p. 322). The scope of a problem refers to the locality of the problem, *i.e.*, how widespread the problem is. The severity of a usability problem, then, refers to the frequency with which the problem occurs, its impact when occurring, and its persistence (Nielsen, 1994b). Several scales are available to rate these problems. Dumas and Redish (1993, p. 324), for example, give a four-level scale with clear reference to the impact on users' tasks:

- level 1 problems prevent users from completing a task;
- level 2 problems significantly slow down the users' performance and frustrate them;

- level 3 problems have a minor effect on usability;
- level 4 problems point to potential upgrades in the future.

The analysis and rating can be done while observing the sessions and making notes, or afterwards with video data analyses by several evaluators. Kjeldskov, Skov, and Stage (2004), for example, present Instant Data Analysis, which can be used to analyze the tests during the test days. In this method, a specific facilitator helps to summarize the notes and findings of the moderator and a note taker at the end of each test day. In their studies, Kjeldskov *et al.* (2004) have been able to detect 85% of the critical usability problems in the system in only 10% of the time required for the corresponding video data analysis.

Communicating the results is a crucial phase in the evaluation process if improvements are desired. The results should not depress the developers but give ideas for making the system even better. The positive findings should therefore also be reported, and the total number of usability problems should be kept to a manageable scale, such as 15–60 problems, focusing only on the most important ones (Molich, Ede, Kaasgaard & Karyukin, 2004).

The results of the analysis should be presented in a report along with redesign proposals. If possible, the test team should test the recommendations with a few users to validate the ideas. Keeping the business goals in mind, and reflecting the proposed changes to these goals also improve the utility of the test reports for the developers (Hornbæk & Frøkjær, 2008). Overall, the results should be communicated in multiple ways to the development team and managers: in written report, verbally in a meeting or in a workshop, and visually with video clips from the test sessions.

International Standard 13407 (1999) gives an example of the contents of a usability evaluation report focusing on formative evaluation, and ISO/IEC 25062 (2006), defines a Common Industry Format dedicated to summative evaluation. The main structures of these report formats are presented in Table 14.2.

Modifications of Usability Testing

The thinking-aloud method is widely used but it is not applicable in all situations, such as in testing systems that are used in critical conditions, or in testing with young children. Some test users also find the method unnatural and distracting. Furthermore, the studies by Ericsson and Simon (1980) show that thinking aloud may slow down the users' performance, so the test users may pay more attention to details that they would neglect in normal use (Rubin & Chisnell, 2008, p. 205). Several alternative methods have therefore been developed for various needs to get into the users' thoughts and experiences when using the assessed systems. This subchapter presents some usability testing methods that have been developed to overcome various problems with the traditional laboratory testing.

Paired-user testing is one way to make the thinking aloud more natural for the users. It involves two users together trying to solve a problem or exploring a system. It has been used for a long time, so it has several names, such as constructive interaction, codiscovery learning, team usability testing, paired-user testing and coparticipation.

Table 14.2 Contents of usability test reports as instructed in ISO 13407 (1999) and ISO/IEC 25062 (2006).

ISO 13407 (Annex B): Formative test	*ISO/IEC 25062: Summative test*
Executive summary	Title page
Product evaluated	Executive summary
Objectives of evaluation	Introduction
Context of use	Full product description
Measurement plan	Test objectives
Users	Method
Methods	Participants
Sequence	Context of product use in the test
Results	Tasks
General	Test facility
Video analysis	Participant's computing environment
User interface design	Test administrator tools
Workflow and process	Experimental design
Training	Procedure
User debriefing	Participant general instructions
User perception questionnaires	Participant task instructions
Recommendations	Usability metrics
Appendices	Effectiveness
	Efficiency
	Satisfaction
	Results
	Data analysis
	Presentation of the results
	Performance results
	Satisfaction results
	Appendices

In paired-user testing, the participants are encouraged to experiment with the studied system, and they are disturbed or interrupted only if the discussion ends. The participants explain their ideas and rationale behind their hypotheses to their partner, so they need to know each other beforehand and have comparable expertise to make an equal and relaxed rapport. The moderator can stay further from the test users in paired-user testing, as the users are engaged in analyzing and exploring the system. We have used the method for example when evaluating televisions, gaming slot machines and office systems and phones.

Peer tutoring is another way to make use of the natural interactions between two users. For example, Höysniemi, Hämäläinen, and Turkki (2003) used peer tutoring to evaluate an interactive computer game with children. Similarly, we used peer tutoring when we evaluated a new educational system for 9–10 years old children (Kantosalo, 2014). We had one child at a time first learning to use the system, and then teaching it to his/her friend. This way, the children could use their own language, and focus on things that they were most interested.

We have used peer tutoring also with adults, for example by including a third party entering the test room during the test. When studying work related systems, this third party has acted as a new trainee, and with a recreational system, the role has been of a relative or a friend asking for an advice in using the evaluated system. This setting has helped in revealing the users' doubts and uncertainties in controlling the system.

Pluralistic usability walkthrough is a usability evaluation method bringing representative users, system designers, and usability experts together to evaluate and discuss on new design ideas (Bias, 1994). The discussion is based on tasks that the participants try to perform with the help of a paper prototype, such as a set of user interface sketches of the system. The participants get copies of the dialogues that they need to perform the given tasks, and instructions to which dialogue to proceed according to their actions.

Documentation or help functions are rarely available at this point, so the system designers usually serve as "living documentation," and answer questions that users indicate they would try to solve with the help of the system documentation. In this way, the users are able to carry on with their tasks, and the designers obtain valuable hints for documentation and further development.

In the original method by Bias, the pluralistic usability walkthrough combines experts doing usability inspections and users commenting on the system. However, we have kept these separate so that the users can be the focus of the sessions. For the same reason, we let the users start the discussion, and only after all the users have commented on the task are the system designers allowed to say which solutions the system supports. The designers usually suggest some new ideas for the system based on the users' comments, and all the participants are welcome to comment these ideas and to generate new ones. (Riihiaho, 2002, 2015)

Also **backtracking analysis** (Akers, Jeffries, Simpson & Winograd, 2012) gathers several users at the same time to do predefined tasks in the same location. During the task performance, the logging is triggered as the user initiates undo or erase functions. After the participants have finished the tasks, they are paired up for retrospective discussions on these logged incidents. First, the other participant comments the logged incidents, when the other one asks the questions prompted by the analysis tool. The answers are audio recorded, and integrated to the logged data for further analysis. After the first participant has gone through all his incidents, the participants switch roles. This way, the amount of material for further analysis is substantially smaller compared to traditional usability tests, as only the critical incidents are recorded and explained (Akers *et al.,* 2012).

Visual walkthrough is a user testing method that has been developed in our research group to get information about users' perceptions and interpretations of the evaluated user interface and its components (Nieminen & Koivunen, 1995). The method can be used to complement a usability test or as a separate method. During a visual walkthrough, the users are not allowed to explore the user interface but to concentrate on one view at a time. At first, the users are asked to tell what they see and notice on the screen. After this general overview, the users are asked to describe what kind of elements, groups and details they notice. The next step is to ask the users to explain what they think the terms and symbols mean, and what kind of functionalities they think the elements provide. After that, a task may be presented, and the users are asked to state their intentions without actually doing the actions (Nieminen & Koivunen, 1995).

We have used the visual walkthrough method also in a modified version that evaluates the utility of a system in addition to its usability. In the modified version, we ask the tests users to mark with different colors in the hard copies the parts that they need and use the most, parts that are rather important, and parts that they never use or that they find useless (Juurmaa, Pitkänen, Riihiaho, Kantola, & Mäkelä, 2013). The participants in this method need to be experts in the domain, so that they are able to assess what information is relevant.

To present the results of the coloring tasks, we combined some user interface elements into blocks, and colored these blocks either with one color if the users' colorings were convergent, or with stripes if the colorings varied a lot. These aggregated block maps showed the most relevant parts as green, whereas the red color indicated blocks that the participants had never used or used only rarely. This coloring method proved to be simple and inexpensive, applicable even with paper and pens. We also found the colored block maps to be valuable tools in communicating the results in a convincing and easily understandable format to the customers (Juurmaa *et al.*, 2013; Riihiaho, 2015).

Informal walkthrough is a mixture of usability testing, observation, and interview. It focuses on evaluating the intuitiveness of a system, so it does not use predefined or specific test tasks. Instead, the test moderator has a list of features that the users are supposed to walk through at their own pace and in their own order, and the ones that require feedback for further development are encircled in the list. As the users explore the system, the moderator marks the visited features with an "*X*" if the users find the feature themselves; with an "*A*" if they find it by accident; and with a "–" if the users do not try out the feature. After the users have explored the system, the moderator checks if some features that require feedback are still not visited. Scenarios, tasks or questions can be used in these situations to guide the users to these features (Riihiaho, 2009, 2015).

Informal walkthrough can be applied both in single-user tests and in multiple-user tests, and it can be used in the laboratory settings as well as in the field. It has been an excellent method, for example, for evaluating gaming slot machines and digital services whose concept is already familiar to the test users.

Contextual walkthrough, then, is a method to evaluate the use of a system in a real use context. It includes elements from traditional usability testing and contextual inquiry. The four main principles of contextual inquiry (Beyer & Holtzblatt, 1998, pp. 37–38), *i.e.*, context, partnership, interpretation, and focus, also apply in contextual walkthrough. The context must be real; the test user and the evaluator should take the roles of a master and an apprentice; the evaluator should interpret what she sees and validate these interpretations with the test user; and there should be a predefined focus in the evaluation.

Contextual walkthrough is a convenient method in evaluating professional systems that cannot be separated from their use contexts, such as call center services. Furthermore, the real use context with realistic tasks gives an excellent opportunity to assess both the usability and utility of the system.

The systems to be evaluated with contextual walkthrough must be systems that the test users are already using or will start using if the system is taken into use. The users need something to try out with their own material, so the system must be on a level of a functional prototype or a running system. The users also need to be experienced enough to take the role of a master showing how the work is done (Riihiaho, 2009, 2015).

Experience clip (Isomursu, Kuutti, & Väinämö, 2004) also utilizes the real-use context. It can evaluate mobile applications and devices outdoors with real users and

Table 14.3 Summary of methods for usability testing and their key information.

Method	*Special focus*	*Level of evaluation*	*Time frame*	*Context*	*Tasks*	*Users per session*	*Level of prototype*
Traditional usability testing		Operation	Test session	Controlled	Predefined	Single	Functional
Backtracking analysis	Automatically logged incidents	Operation	Test session	Controlled	Predefined	Several	Functional
Paired-user testing		Operation	Test session	Controlled or real use context	Predefined	Pair	Functional
Peer tutoring	Learning	Operation	Test session	Controlled	"*Teach the tutee to use the system*"	Pair or more	Functional
Pluralistic usability walkthrough		Operation	Test session	Controlled	Predefined	Two or more	Paper prototype
Modified visual walkthrough	Relevance of user interface elements	Operation, utility	Intended use	Controlled	No tasks performed	Single	Paper prototypes or printouts
Informal walkthrough	Intuitiveness	Operation, utility, value	Test session	Controlled or real use context	Free and some predefined scenarios	Single or more	Functional
Contextual walkthrough	Feasibility of work-related systems	Operation, utility, value	Test session	Real use context	Real tasks from the context	Single or more	Functional
Experience Clip	One user testing and other recording	Operation, utility, value	Brief use on own	Real use context	Real tasks	Pair	Functional

without evaluators observing the test users all the time. In this method, pairs of users are invited into the study from the passers by. The other participant gets a mobile phone with the evaluated application, and the other one gets a mobile phone with a video shooting capability. The latter one then takes video clips as the first participant uses the application. When returning the equipment, the participants select the clips to go through, and briefly describe what they have done with the application, and how the application has worked (Isomursu *et al.,* 2004).

These examples of usability testing methods show that the traditional test setting in specific usability laboratories can be modified in many ways according to the goals of the evaluations. In formative testing that does not aim for statistical significance but for rich feedback from the users, the methods, settings, and prototypes can be modified between tests and even during a test session to maximize the users' comments about new ideas and redesign proposals.

Every method has its strengths and requirements, so it is recommended to combine a set of methods in each evaluation process. Table 14.3 summarizes briefly the methods presented in this chapter by presenting the level in which the method permits users to assess the system (operation, utility or value), the time interval that is studied, the environment in which the method is intended to be used, the source of the test tasks, number of intended test users in one test session, and the level of prototype required (paper prototype, functional prototype, or working system).

The summary shows that all the methods focus on only a brief test session or a brief exploration. Therefore, evaluations must be repeated or the users must be given a chance to try out the system already before the tests if longitudinal studies are needed.

Most of the methods also require something that the users can actually try out, *i.e.*, a functional prototype or even a working system. Only the pluralistic usability walkthrough and visual walkthrough methods are designed for paper prototypes. Pluralistic usability walkthrough gives room for group discussions, so it is a good method in the early phases of system development to assess new ideas.

Furthermore, the summary shows that quite few of the methods support the assessment of other issues besides usability of the system. In practice, however, most of the methods can be modified to support assessing other attributes as well. If the test users are allowed to explore the system on their own, may select their own tasks, or can use realistic and personal data, they are more competent to assess also the usefulness of the system, and even the value of the system in the users' daily life and social contexts.

References

Akers, D., Jeffries, R., Simpson, M., & Winograd, T. (2012). Backtracking events as indicators of usability problems in creation-oriented applications. *ACM Transactions on Computer-Human Interaction, 19*(2), 16:1–16:40. doi:10.1145/2240156.2240164

Bangor, A., Kortum, P. T., & Miller, J. T. (2008). An Empirical Evaluation of the System Usability Scale. *International Journal of Human-Computer Interaction, 24*(6), 574–594. doi:10.1080/10447310802205776

Beyer, H. R., & Holtzblatt, K. (1995). Apprenticing with the customer. *Communications of the ACM, 38*(5), 45–52. doi:10.1145/203356.203365

Beyer, H. R., & Holtzblatt, K. (1998). *Contextual design: Defining customer-centered systems.* San Francisco, CA: Morgan Kaufmann.

Bias, R. (1994). The pluralistic usability walkthrough: Coordinated empathies. In J. Nielsen & R. L. Mack (Eds.), *Usability inspection methods* (pp. 63–76). New York, NY: John Wiley & Sons, Inc.

Bødker, S., & Madsen, K. (1998). Context: An active choice in usability work. *interactions, 5*(4), 17–25. doi:10.1145/278465.278469

Brooke, J. (1996). SUS: A "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester & I. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189–194). London: Taylor & Francis.

Brooke, J. (2013). SUS: A retrospective. *Journal of Usability Studies, 8*(2), 29–40.

Clemmensen, T., Hertzum, M., Hornbæk, K., Shi, Q., & Yammiyavar, P. (2009). Cultural cognition in usability evaluation. *Interacting with Computers, 21*(3), 212–220. doi:10.1016/j.intcom.2009.05.003.

Dumas, J. S., & Loring, B. A. (2008). *Moderating usability tests: Principles and practices for interacting*. Amsterdam: Elsevier.

Dumas, J. S., & Redish, J. C. (1993). *A practical guide to usability testing*. Norwood, New Jersey: Ablex.

Ericsson, K. A., & Simon, H. A. (1980). Verbal reports as data. *Psychological Review, 87*(3), 215–251. doi:10.1037/0033-295X.87.3.215

Gomoll, K. (1990). Some techniques for observing users. In B. Laurel & S. J. Mountford (Eds.), *The art of human-computer interface design* (pp. 85–90). Boston, MA: Addison-Wesley.

Hassenzahl, M., & Sandweg, N. (2004). From mental effort to perceived usability: Transforming experiences into summary assessments. In *CHI '04 Extended Abstracts on Human Factors in Computing Systems* (pp. 1283–1286). New York, NY: ACM. doi:10.1145/985921.986044

Held, J. E., & Biers, D. W. (1992). Software usability testing: Do evaluator intervention and task structure make any difference? In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 36*(16), 1215–1219. doi:10.1177/154193129203601607.

Hewett, T. T. (1986). The role of iterative evaluation in designing systems for usability. In *Proceedings of the Second Conference of the British Computer Society, human computer interaction specialist group on people and computers: Designing for usability* (pp. 196–214). New York, NY: Cambridge University Press.

Hornbæk, K. (2006). Current practice in measuring usability: Challenges to usability studies and research. *International Journal of Human-Computer Studies, 64*(2), 79–102. doi:10.1016/j.ijhcs.2005.06.002

Hornbæk, K., & Frøkjær, E. (2008). Making use of business goals in usability evaluation: An experiment with novice evaluators. In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* (pp. 903–912). New York, NY: ACM. doi:10.1145/1357054.1357197

Hornbæk, K., & Law, E.L-C. (2007). Meta-analysis of correlations among usability measures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 617–626). New York, NY: ACM. doi:10.1145/1240624.1240722

Höysniemi, J., Hämäläinen, P., & Turkki, L. (2003). Using peer tutoring in evaluating the usability of a physically interactive computer game with children. *Interacting with Computers, 15* (2), 203–225. doi:10.1016/S0953-5438(03)00008-0

Hwang, W., & Salvendy, G. (2010). Number of people required for usability evaluation: The 10 ± 2 rule. *Communications of the ACM, 53*(5), 130–133. doi:10.1145/1735223.1735255

ISO 13407 (1999) SFS-EN ISO 13407: Vuorovaikutteisten järjestelmien käyttäjäkeskeinen suunnitteluprosessi. *Human-centred design processes for interactive systems.* Suomen Standardoimisliitto SFS ry Helsinki, and European Committee for Standardization, Brussels.

ISO 9241-11 (1998) SFS-EN ISO 9241-11: Näyttöpäätteillä tehtävän toimistotyön ergonomiset vaatimukset. Osa 11: Käytettävyyden määrittely ja arviointi. *Ergonomic requirements* *for office work with visual display terminals (VDTs). Part 11: Guidance on usability.* Suomen Standardoimisliitto SFS ry Helsinki, and European Committee for Standardization, Brussels.

ISO/IEC 25062 (2006). *Software engineering—Software product Quality Requirements and Evaluation (SQuaRE)—Common Industry Format (CIF) for usability test reports.* International Organization for Standardization and International Electrotechnical Commission, Switzerland.

Isomursu, M., Kuutti, K., & Väinämö, S. (2004). Experience clip: Method for user participation and evaluation of mobile concepts. In *Proceedings of the eighth conference on Participatory design: Artful integration: Interweaving media, materials and practices* (pp. 83–92). New York, NY: ACM. doi:10.1145/1011870.1011881

Juurmaa, K., Pitkänen, J., Riihiaho, S., Kantola, T., & Mäkelä, J. (2013). Visual walkthrough as a tool for utility assessment in a usability test. In *Proceedings of the 27th International BCS Human Computer Interaction Conference* (pp. 1–6). http://ewic.bcs.org/upload/pdf/ewic_hci13_short_paper11.pdf

Kantosalo, A. (2014). *Usability evaluation with children—case: Poetry machine* (Master's thesis, Aalto University, Espoo). Retrieved from http://urn.fi/URN:NBN:fi:aalto-201412033119

Kantosalo, A., & Riihiaho, S. (2014). Let's play the feedback game. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction.* (pp. 943–946). New York, NY: ACM. doi:10.1145/2639189.2670246

Kim, H., Kim, J., Lee, Y., Chae, M., & Choi, Y. (2002). An empirical study of the use contexts and usability problems in mobile Internet. In *Proceedings of the 35th Annual Hawaii International Conference on System Sciences* (pp. 1767–1776). doi:10.1109/HICSS.2002.994090

Kjeldskov, J., Skov, M. B., & Stage, J. (2004). Instant data analysis: Conducting usability evaluations in a day. In *Proceedings of the third Nordic conference on human-computer interaction* (pp. 233–240). New York, NY: ACM. doi:10.1145/1028014.1028050

Lewis, J. R. (1994). Sample sizes for usability studies: Additional considerations. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 36*(2), 368–378. doi:10.1177/001872089403600215

Lewis, J. R. (2001). Evaluation of procedures for adjusting problem-discovery rates estimated from small samples. *International Journal of Human-Computer Studies, 13*(4), 445–479. doi:10.1207/S15327590IJHC1304_06

Lindgaard, G., & Chattratichart, J. (2007). Usability testing: What have we overlooked? In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 1415–1424). New York, NY: ACM. doi:10.1145/1240624.1240839.

Maguire, M. (2001). Context of use within usability activities. *International Journal of Human-Computer Studies, 55*(4), 453–483. doi:10.1006/ijhc.2001.0486

Markopoulos, P., & Bekker, M. (2003). On the assessment of usability testing methods for children. *Interacting with Computers, 15*(2), 227–243. doi:10.1016/S0953-5438(03)00009-2

McDonald, S., Monahan, K., & Cockton, G. (2006). Modified contextual design as a field evaluation method. In *Proceedings of the 4th Nordic Conference on Human-Computer Interaction: Changing Roles* (pp. 437–440). New York, NY: ACM. doi:10.1145/1182475.1182531

Molich, R., Ede, M. R., Kaasgaard, K., & Karyukin, B. (2004). Comparative usability evaluation. *Behaviour and Information Technology, 23*(1), 65–74. doi:10.1080/0144929032000173951

Nielsen, J. (1993). *Usability engineering.* Boston, MA: Academic Press.

Nielsen, J. (1994a). Estimating the number of subjects needed for a thinking aloud test. *International Journal of Human-Computer Studies, 41*(3), 385–397. doi:10.1006/ijhc.1994.1065

Nielsen, J. (1994b). Heuristic evaluation. In J. Nielsen & R. L. Mack (Eds.), *Usability inspection methods* (pp. 25–62). New York, NY: John Wiley & Sons, Inc.

Nieminen, M., & Koivunen, M.-R. (1995). Visual walkthrough. In *Adjunct Proceedings of HCI'95: People and Computers. The 10th International BCS Human Computer Interaction Conference* (pp. 86–89). Huddersfield: University of Huddersfield.

Read, J. C., & MacFarlane, S. (2006). Using the fun toolkit and other survey methods to gather opinions in child computer interaction. In *Proceedings of the 2006 Conference on Interaction Design and Children* (pp. 81–88). New York, NY: ACM. doi:10.1145/1139073.1139096

Riihiaho, S. (2002). The pluralistic usability walk-through method. *Ergonomics in Design: The Quarterly of Human Factors Applications, 10*(3), 23–27. doi:10.1177/106480460201000306

Riihiaho, S. (2009). User testing when test tasks are not appropriate. In *European Conference on Cognitive Ergonomics: Designing beyond the Product—Understanding Activity and User Experience in Ubiquitous Environments* [Proceedings of ECCE 2009] (pp. 228–235). Espoo: VTT Technical Research Centre of Finland. Retrieved from http://dl.acm.org/citation.cfm?id=1690537&CFID=541456289&CFTOKEN=10679551

Riihiaho, S. (2015). *Experiences with usability testing: Effects of thinking aloud and moderator presence* (Doctoral dissertation, Aalto University, Helsinki). Retrieved from http://urn.fi/URN:ISBN:978-952-60-6227-3

Riihiaho, S., Nieminen, M., Westman, S., Addams-Moring, R., & Katainen, J. (2015). Procuring usability: Experiences of usability testing in tender evaluation. In *Scandinavian Conference on Information Systems* (pp. 108–120). Cham, Switzerland: Springer International Publishing. doi:10.1007/978-3-319-21783-3_8

Rubin, J., & Chisnell, D. (2008). *Handbook of usability testing: How to plan design and conduct effective tests* (2nd ed.). Indianapolis, IN: John Wiley & Sons, Inc.

Sauer, J., & Sonderegger, A. (2009). The influence of prototype fidelity and aesthetics of design in usability tests: Effects on user behaviour, subjective evaluation and emotion. *Applied Ergonomics, 40*(4), 670–677. doi:10.1016/j.apergo.2008.06.006

Sauro, J., & Lewis, J. R. (2009). Correlations among prototypical usability metrics: Evidence for the construct of usability. In *Proceedings of the 27th International Conference on Human Factors in Computing Systems* (pp. 1609–1618). New York, NY: ACM. doi:10.1145/1518701.1518947

Sauro, J., & Lewis, J. R. (2011). When designing usability questionnaires, does it hurt to be positive? In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2215–2224). New York, NY: ACM. doi:10.1145/1978942.1979266

Schrier, J. R. (1992). Reducing stress associated with participating in a usability test. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting. 36*(16), 1210–1214. doi:10.1177/154193129203601606.

Schulte-Mecklenbeck, M., & Huber, O. (2003). Information search in the laboratory and on the Web: With or without an experimenter. *Behavior Research Methods, Instruments, and Computers, 35*(2), 227–235. doi:10.3758/BF03202545.

Skov, M. B., & Stage, J. (2012). Training software developers and designers to conduct usability evaluations. *Behaviour and Information Technology, 31*(4), 425–435. doi:10.1080/01449290903398208.

Sonderegger, A., & Sauer, J. (2009). The influence of laboratory set-up in usability tests: effects on user performance, subjective ratings and physiological measures. *Ergonomics, 52*(11), 1350–1361. doi:10.1080/00140130903067797

Tohidi, M., Buxton, W., Baecker, R., & Sellen, A. (2006a). Getting the right design and the design right. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1243–1252). New York, NY: ACM. doi:10.1145/1124772.1124960

Tohidi, M., Buxton, W., Baecker, R., & Sellen, A. (2006b). User sketches: A quick, inexpensive, and effective way to elicit more reflective user feedback. In *Proceedings of the 4th Nordic Conference on Human-computer Interaction* (pp. 105–114). New York, NY: ACM. doi:10.1145/1182475.1182487

Virzi, R. A. (1992). Refining the test phase of usability evaluation: How many subjects is enough? *Human Factors: The Journal of the Human Factors and Ergonomics Society, 34*(4), 457–468. doi:10.1177/001872089203400407

Yeo, A. (2000). Are usability assessment techniques reliable in non-Western cultures? *The Electronic Journal of Information Systems in Developing Countries, 3*(1), 1–21.

15

Remote Usability Testing

John Black and Marc Abrams

Introduction

The Army's call for technology

In 2013, the US Army put out a request for technology to fill a gap they perceived in existing tools used to build or buy complex systems with human interfaces. In their experience, too many software systems are paid for and accepted by the Army, and other US federal agencies, with user interfaces that are just too hard to use. This results in excessive training costs, errors—sometimes fatal—and expensive rewrites or rejections. While it was possible for experts, the request said, to test new software systems before final acceptance at great time and expense, what the Army needed was a tool that let the average purchasing agent in the federal government determine if a new system would be easy to use, quickly, and at a reasonable cost. Harmonia heard this call, proposed a solution, and was granted a contract to create new technology for testing the usability of new systems interfaces before final acceptance by the Army.

Harmonia's GUITAR solution

Harmonia worked with the Army for about 3 years to build a product named GUITAR, an acronym for Graphical User Interface, Test, Assess, and Remediate, which enables this kind of semiautomatic, inexpensive, remote usability testing. In this chapter, I use our experience building that product as a running example while discussing remote user testing. It is both a software as a service (SaaS) website and a tool for doing remote user testing. One of the authors of this article was the lead developer of this project.

Remote usability testing

Remote user testing (www.usability.gov) is first of all *user* testing. But it is a type of user testing that is done where the researcher is separated from the test participant. Usability testing itself is treated in another chapter of this handbook; here I focus on the need for, advantages of, and the practical challenges of doing such testing *remotely*.

The Wiley Handbook of Human Computer Interaction, Volume 1, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

User versus usability testing

Additionally, the terms "user testing" and "usability testing" are used in various ways in the literature (Anderson, 2011). Our interest in this chapter is on the effect of the remoteness between the tester and the tested subject, so I will use those two terms as follows. User testing will be used as the broader category, with usability testing used to mean a particular kind of user testing that focuses on the ease of use a user experiences when using a system interface.

So for us, usability testing is a *type* of user testing, because user testing may include other research into a user's experience. You could, for example, test a user's *familiarity* with a system, as a part of user testing. In this chapter, I am generally concerned with user testing for usability. Thus, the terms are for us practically interchangeable. But, in any case, I am concerned with *remote* user and usability testing, where test participants and investigators are separated from each other in both space and time.

Moderated versus unmoderated remote user testing

This brings up another distinction. Remote user testing is often divided broadly into two kinds: moderated and unmoderated (Varga, 2011). In *moderated* remote testing there is a separation of space between the researcher and the test participant—they are not in the same room. But the testing still happens simultaneously for the researcher and the test participant. They are in direct communication with each other in real time, for example, in a telephone conversation. Now just as a telephone call differs from a personal conversation, the physical distance between the researcher and test participants leads to differences in the testing that takes place. But while this is *remote*, in a sense, it is in many respects closer to face-to-face usability testing than not. There is the ordinary conversational give and take. The researcher can respond to what a test participant says in real time and alter the direction of the interview. But the need for the researcher and test participant to work simultaneously results in scheduling and coordination costs that are similar to those found in face-to-face testing.

With *unmoderated* remote usability testing, the researcher is separated from the test participant by both space and by time and there is no real-time communication between them. In this type of remote user testing, the test participant interacts with a target system interface while another system records their physical and verbal behavior, and other measurable responses, for later analysis.

In the current chapter, I will focus primarily on this second type, unmoderated remote user testing. This is the case for the GUITAR project work at Harmonia. We see *unmoderated remote user testing* as the most promising way to meet the requirements of the Army to innovate technology that will enable even nonexpert buyers to evaluate a new systems interface, quickly and cheaply. I will make the reasons for this clear in the rest of this chapter.

Kinds of user research

In this chapter, I am going to limit the discussion to user *testing*. There are many methods of doing user research (Rochanayon, 2015). The concept of user *testing* involves giving actual users an opportunity to use the product and observing and measuring the ease with which they can accomplish their intentions. Ideally, actual users will be observed and measured while using a target system. This can be distinguished from having an expert evaluate a system by comparing it to known principles of usability, generally referred to as *heuristic evaluation*. It is also distinguished from finding test subjects who are similar enough to an actual user testing a system in a more or less simulated situation. Additionally, it is not like those testing methods that subject an automated simulation of a human being to a systems interface to detect characteristics that are problematic.

Justifying the limitation of remote user testing

The separation in space and time and the asynchronous and indirect communication between the researcher and the test participant in unmoderated remote user testing could lead to a degradation of the quality of data that can be collected. Therefore, there must be a real need for, or advantages in, using this method of testing that would justify and offset this loss of quality. I will discuss the need for and advantages of unmoderated remote user testing next.

The Need for Remote Testing

The best user test

The need for remote unmoderated user testing becomes clear when reflecting on the following. The ultimate value of user testing could be best achieved if it was possible to test actual users, in real time, during an instance of actual use of a system interface, without any disruption of their efforts. Then the data gained would be as good it gets. Every other testing situation can be judged relative to this.

Circumstances that require remote testing

In some cases, there are circumstances where remote usability testing is clearly necessary or desirable (Bolt & Tulathimutte, 2010). Here are some examples.

Laboratory could be prohibitively expensive It may be prohibitively expensive to bring actual users into the laboratory for testing. And it may be very difficult to find surrogates with characteristics close enough to ensure valid results. This could be because the users are distributed over a wide geographical area. For example, if your system is used by war fighters deployed to theaters all over the globe, then the cost of bringing them back to the laboratory could be excessive. Or the actual users may not be willing to spend the time to come to a laboratory for testing. This could be a particular problem, for example, when testing software designed for users such as lawyers, doctors, entertainers, or sports stars.

Context of use is expensive to simulate The laboratory setting may alter the conditions of use so much that testing would be invalid. This is sometimes overcome by building elaborate and expensive simulations of the conditions under which a system is used. Airline cockpits, for example, have been simulated for training pilots safely. Such simulations have been used for testing additions or changes to system interfaces used to pilot an aircraft. But in the absence of the multimillion dollar budgets needed to build a simulation, the laboratory is often too different from the actual context of use to produce valid test results. As a more mundane example, a directional aid program for drivers needs to be used in an automobile while people are looking for a destination. Bringing users into the laboratory, where there is no traffic and the car is not moving, would change the test significantly in ways that could prevent a valid test being done at all.

Time aware research The use of all system interfaces are to some extent "time aware" and would benefit if testing could be done at specific times (Bolt & Tulathimutte, 2010). Some require it. If your user research requires testing a system just at the time in which certain events take place and there is no way to simulate or duplicate these events, you must find a way to record the users experience at the time that it occurs. This can be done with remote user testing. For example, emergency warning systems, designed to interrupt and alert the user to an imminent or active emergency, would be difficult to test in a laboratory setting, although it could be done with some elaborate stage setting and preparation. Even then, it might not be ethical to try to stage a situation that put someone in the mindset of imminent danger.

Advantages of Remote Testing

Test participants are located remotely

The process of locating users, arranging for laboratory time, and making sure all the equipment and all the people come together simultaneously in the same place can be very difficult and very expensive. With remote unmoderated user testing, test participants can be located practically anywhere. Using recording devices to gather usage data that can be transmitted back to the researcher means that the testing can take place at any time it is convenient for the participant. These are some of the advantages of remote unmoderated user testing.

Live recruiting

It is the wide availability of the Internet, which can be used to locate and recruit test participants, that provides another big advantage. Such "live recruiting" is very valuable. It allows you to find your test participants in their natural environment doing whatever they are normally doing. Although there are issues, the use of surveys and other questionnaires can be used to select demographic or psychographic groups of participants that represent target populations. Sign-up forms, identification systems, and other means of vetting can be used to establish identity and characteristics of a potential test participant. There are, in addition, commercial entities that will perform these recruiting and vetting services for a fee.

No moderator means no human bias

Because remote unmoderated user testing is done automatically, testing will not suffer from any kind of bias. Every participant will be treated the same.

Examples of Remote Testing

Commercial products

There are many commercial products available that allow researchers to perform remote unmoderated user testing. A list of them is maintained online at the Neilson Group's web page (Nielsen Norman Group, 2014). These products vary widely in their range of recording technologies, pricing structure, participant recruiting, and other features. They are evolving rapidly and will be mentioned but not exhaustively described here.

Embedded popup surveys

One of the simplest examples of remote usability testing is the popup survey embedded into many web products. These surveys run remotely and come very close to testing real users of an actual system in the ordinary context of use. However, as they generally follow an instance of interaction, they are not an instantaneous measurement. That is, they happen following the actual use. Thus, they are dependent, in addition to everything else, on the accuracy of the memory of the test participants. Furthermore, as self reports, they suffer from discrepancies between opinion and behavior.

Quantified self

There is a movement called "quantified self" whose adherents record, on a continuous basis, many kinds of data about themselves. These data are later subjected to analysis to look for what insights can be obtained. This promises to provide a whole new source of data for remote user testing. If you can find an overlap between the data necessary to the goals of the research that you want to accomplish and the kinds of data that are already being collected by quantified self participants, then it may be possible to incorporate such data into your research with very little effort.

Other embedded tools

Here are several other examples of remote user testing, some of which are not often thought of as user research.

The "black box" data recorder devices that are embedded into airplanes and trains to record the last minutes of user behavior and systems events before an accident occurs are an example of unmoderated remote user testing.

Many software products now contain an automatic usage data recorder embedded in them that runs continuously and transmits data back to the developers of the product to allow them to make improvements and other changes.

Harmonia's remote user testing tool

Harmonia's answer to the call of the US Army is call GUITAR, which stands for Graphical User Interface Test, Assess, and Remediate. It is an example of a tool that can be used to perform remote unmoderated user testing. The qualitative recording of user interactions is handled by an extensible set of components that can be delivered over the Internet and operated by users themselves to record their experience trying out functions in new software systems. This data is then transferred back to the central server for later analysis and coding by users.

Limitations of Remote Testing

Converting expert researchers from manual methods

One of the major problems we found in commercializing GUITAR was that the experts who have been in charge of evaluating new software and systems are reluctant to change their manual methods. In fact, they do not see the problem with what they are doing. It is only when you confront them with the large backlog of untested software, or the large number of systems that don't ever undergo any testing, that they begin to see the value in an automated remote-testing solution.

Even then, their first inclination is to ask how they can use the new tool to do *exactly* what they have been doing manually for so long. This is neither always possible nor desirable.

But the program and project managers, and other personnel who might use the product in the absence of an available expert, are themselves rarely knowledgeable about what it takes to evaluate the usability of a product.

No control over participants

In unmoderated remote testing, you cannot vary the interview depending on the results that are provided by the participant. You can't get participants back on track when they have gone astray. Participants may not do what they were instructed to do and either you won't ever know, or if you found out later, you may have no choice but to throw out the results obtained. Without the presence of the test participant in the room together with the researcher, without the richness provided by the test participant's facial expressions, body language, and vocal tones, the researchers lose the opportunity to empathize fully with what the test participant is feeling during the test. As a result, there is no opportunity to discover and pursue unexpected human reactions to the test.

Massive amounts of data

Because you are not in direct control of the remote recording devices that you have selected, it is possible that far more data than is necessary is collected. All data must be scanned for relevance. It must be observed for issues revealed. Until automated systems are built, perhaps using deep-learning techniques that can scan recordings and detect usability problems, remotely collected data will still need a human researcher to review and code issues. And this can be very time consuming. In personal interviews with Army personnel, the figure that was claimed was ten times the length of each recording.

Tradeoff between reach and certainty

Unless your test participants evince high fidelity to the actual users targeted and the typical context of use, your testing could produce meaningless results. Remote user testing on the one hand can make it easier to find people and situations that are close to ideal but it also presents a problem. It is harder for you to verify and prove that the testing was done with these high fidelity subjects in a realistic context. So that is the double-edged sword of remote unmoderated user testing: you can potentially reach a real user, or someone very similar, using the live product in an actual situation, but it is much harder to know and prove that you have done so.

Remote Testing Workflow

Formative versus summative

Testing can be done at many different times during the process of creation, deployment, maintenance, and revision of existing systems. However, historically, there are two *main* points in time in the software or systems lifecycle that user testing can be used.

First, it may be used during the creation of a new product, at many points in the design, prototyping, and implementation of the product, before it has been delivered, to increase the chance that the final product will be usable. This is called *formative* testing. It has been shown, in studies of testing for functional defects, that the earlier a defect is found, the less expensive it will be to fix it.

The second point in time that products are tested is upon final delivery to the customer. This is called *summative* testing, and takes place at the summation of a software development effort. Testing done at this time could be used in final acceptance or rejection of the product. It sets up a final hurdle that must be leaped before the product is paid for. Knowing this test is coming can be used during development to justify the expense of maintaining formative testing.

GUITAR supports both of these types of testing. In formative testing, design images of proposed screens, for example, can be presented to remote test participant along with questions about how easy it would be to use those screens if they are implemented.

Postrelease testing

Finally, even after the software is released, usability testing should be performed on a regular basis. As the context of use and the demographics of the users may change, usability testing may need to be done on a regular basis to maintain the same quality as the product is deployed.

Comparison of two existing products

Often software is not going to be developed to order. There may already exist one or more products that could be used. If there is more than one, it would be advantageous to be able to evaluate both products to determine, for equal functional goals, which of the two or more prospects was the most usable. This is what is called the "Evaluation of Alternatives" process by the federal government. The GUITAR product was developed with this goal in mind. Once user test plans have been designed, testing sessions can be initiated that will ensure that the same test is administered to each user for each of the products of interest. The results can then be compared.

Using GUITAR

A website for remote unmoderated user testing

GUITAR, is a tool developed by Harmonia Holdings Group, LLC, under a Small Business Innovation Research (SBIR) grant. It is both a software as a service (SaaS) website and a tool for doing remote user testing. One of the authors of this chapter is the lead developer of this project. Figure 15.1 shows the landing screen, which may be accessed at https://guitarux.com.

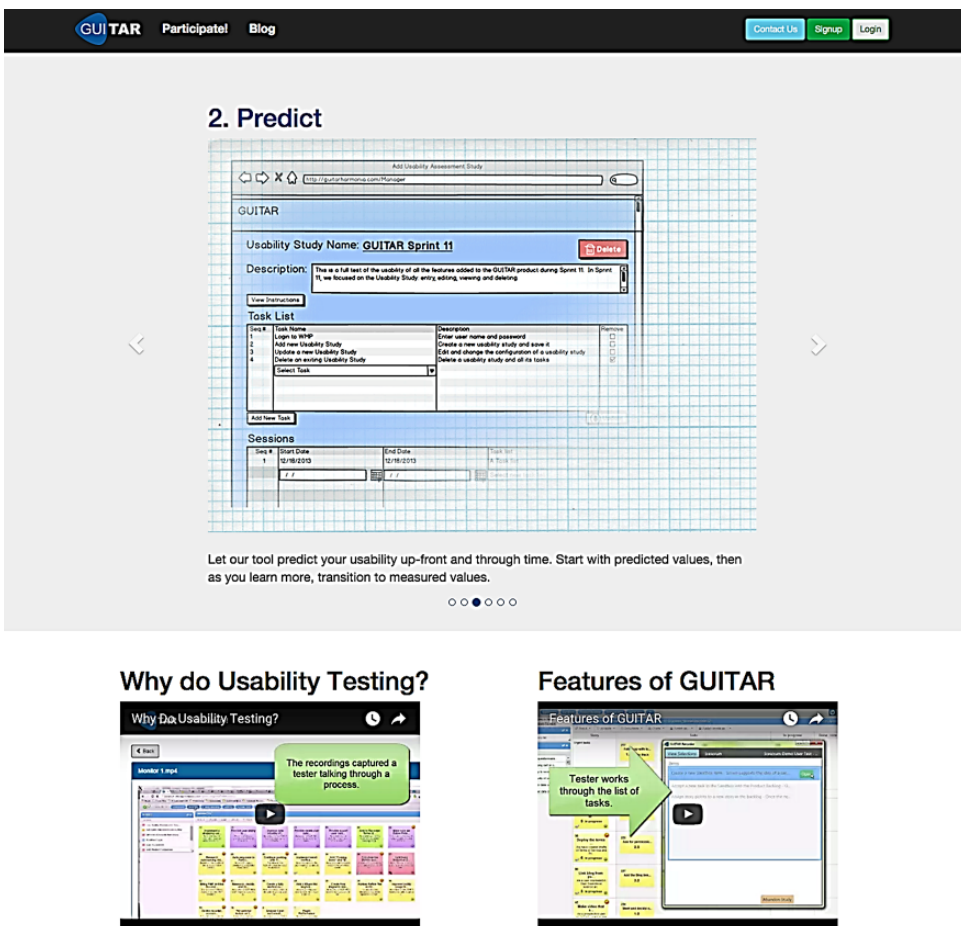


Figure 15.1 Landing page for GUITAR at https://guitarux.com

Managing remote testing projects The process of subjecting a new product under development to usability testing is a complex operation. Harmonia's GUITAR product, as well as others, has features that allow users to keep track of the tests that have been performed on each successive version.

To facilitate this bookkeeping function, GUITAR lets you enter and store the product, the versions, the test plans, the testing sessions, and all the results that are collected. Tests should be repeated as closely as possible, so that the statistics gathered are meaningful.

Test plan development and management The main user study design page, shown in Figure 15.2, lets the test designer give the study a unique name, define the target product and product version under test, describe the study, and choose the study type. Instructions can be provided to guide the test participant during the execution of the study.

A test plan, which is a list of tasks to be performed that can include images of mock-ups of a product yet to be built, along with survey questions to be asked after the completion of each task, affords the test designer a rich array of both qualitative and quantitative data points to collect in an organized and repeatable way.

The researcher can also select from an extendible list of recording devices. This includes video of the participant, screen recordings, mouse movements and clicks, keystrokes, and any other kind of recording device that the researcher can provide a plugin to the software to control.

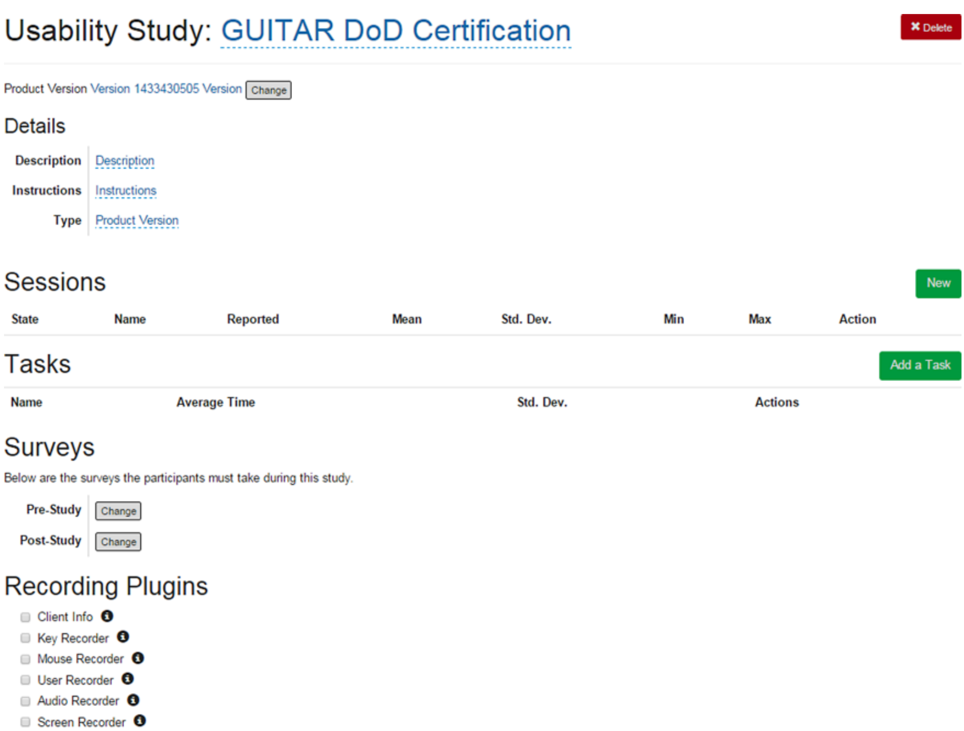


Figure 15.2 Top level usability study design and setup page.

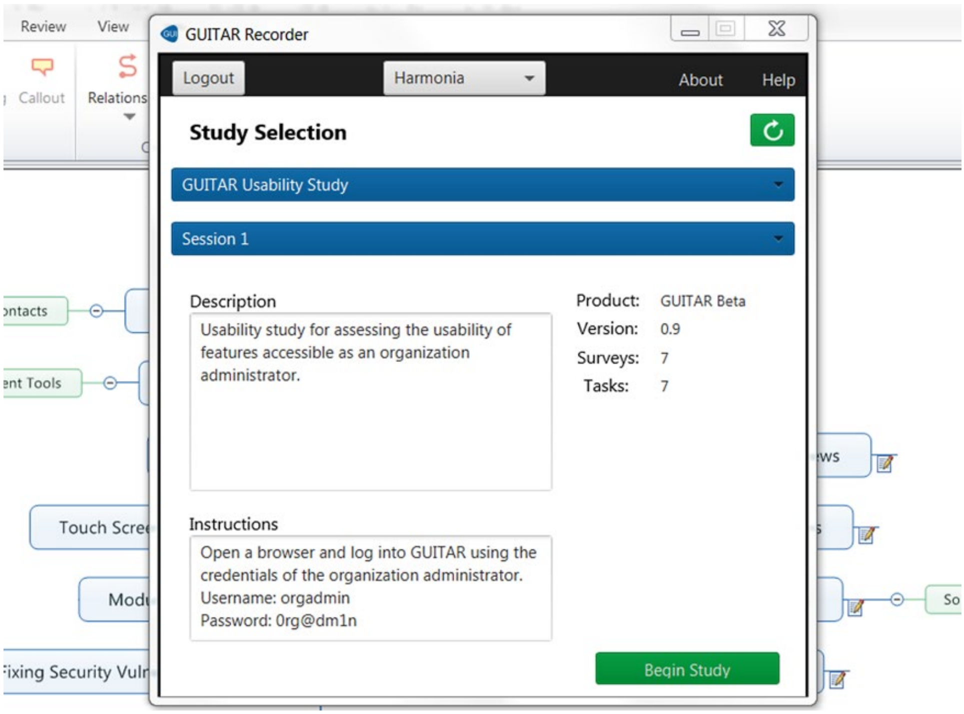


Figure 15.3 Test participant workstation data recorder application.

Data recording and collection The GUITAR data recorder presents the study designer's instructions to the test participants and affords them control over the execution of the various steps of the test. Participants can select from different studies and sessions within those studies, as can be seen in Figure 15.3, and then proceed to call up the individual tasks one by one until they have completed all of the steps of the test plan. Both before the entire test and after the whole thing has been completed, the participants can be presented with sets of survey questions. Smaller sets of survey questions can also be asked following every task that the participant is asked to perform.

The promise of remote unmoderated user testing is in the potential for new kinds of recording technology to become available that will collect even more meaningful data. GUITAR is designed to allow for new recording devices to be plugged into the recorder.

Kinds of data collected Almost any kind of data that can be measured can be collected remotely. And technological advancements continue to add new ways of recording data all the time. GUITAR is built to allow for additional recording devices, using a "plugin" architecture, as shown in Figure 15.4. Among the most common kinds of data that are being collected in the current set of tools that are available at the time this was written include:

- video of the participants;
- video of the participants' screen or other user interface;
- mouse movements and button clicks;
- other input device's motion and clicks;

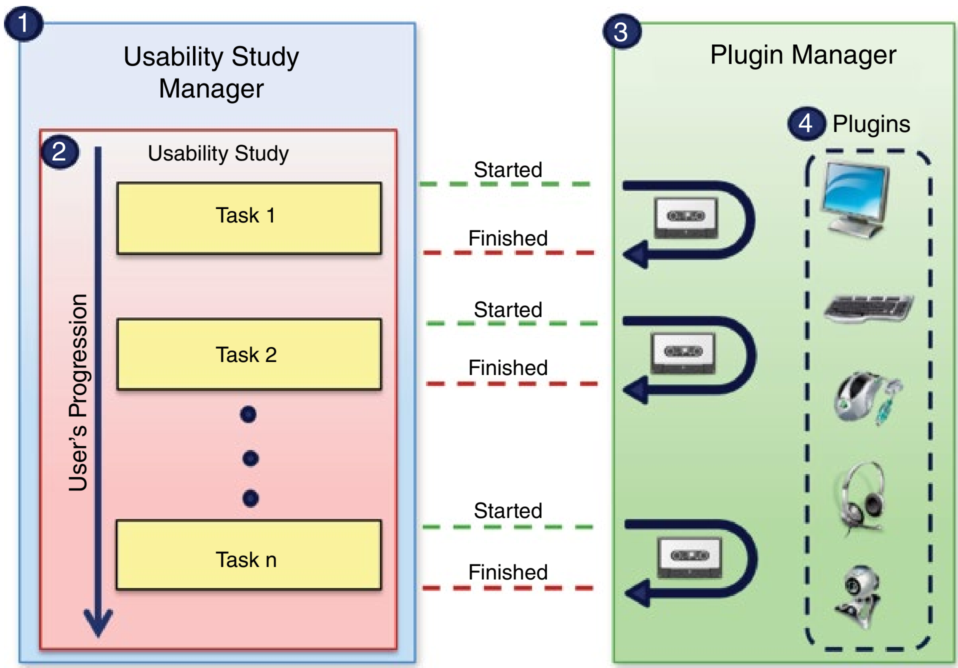


Figure 15.4 GUITAR can be extended with additional recording device plugins.

- keyboard keystrokes;
- audio recordings of all verbal reports;
- eye-tracking data;
- physiological measures such as heart rate.

As Figure 15.5 shows, GUITAR takes these different kinds of data and packages them all in the standard MP4 format, if possible.

We then reconstruct the timeline for all this recorded data to create a synchronized playback in a media player.

Data recording media player and issue coder The GUITAR media player and coding utility will display videos, audio recordings, screen images, mouse clicks, and other recorded media all synchronized together in the Observer component, as can be seen in Figure 15.6. This allows you to use computer assisted qualitative data analysis software (CAQDAS) on each participant's test result. All such data is then indexed per the issue or observation made. Later, when reporting on these issues, the links to the specific location in the media can be included. Stakeholders interested in the issue reported can drill down and view the actual test recording to verify that the situation was, in fact, coded properly.

Create and store surveys While you will read about the use of surveys in the chapter on usability testing in this handbook, I mention them here because it is common practice to use a survey to gather opinions from test participants about the usability of the system. In Harmonia's GUITAR product, for example, we include the capability to create a wide range of possible surveys to proceed or follow the test plan. We also include standard

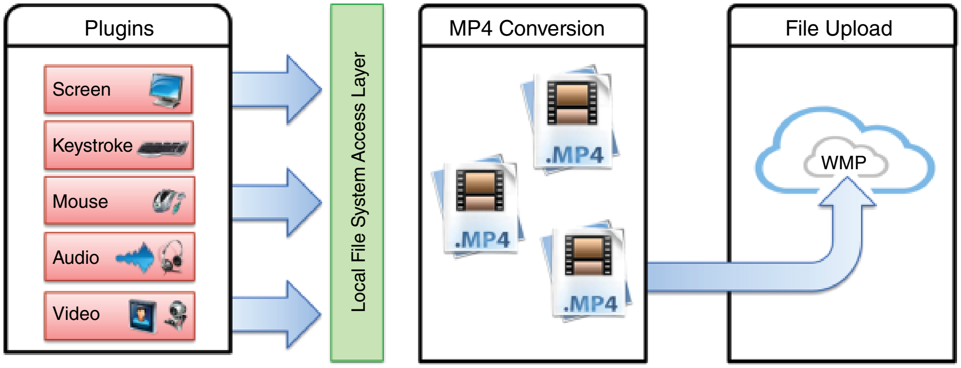


Figure 15.5 Recordings are packaged into standard container formats.

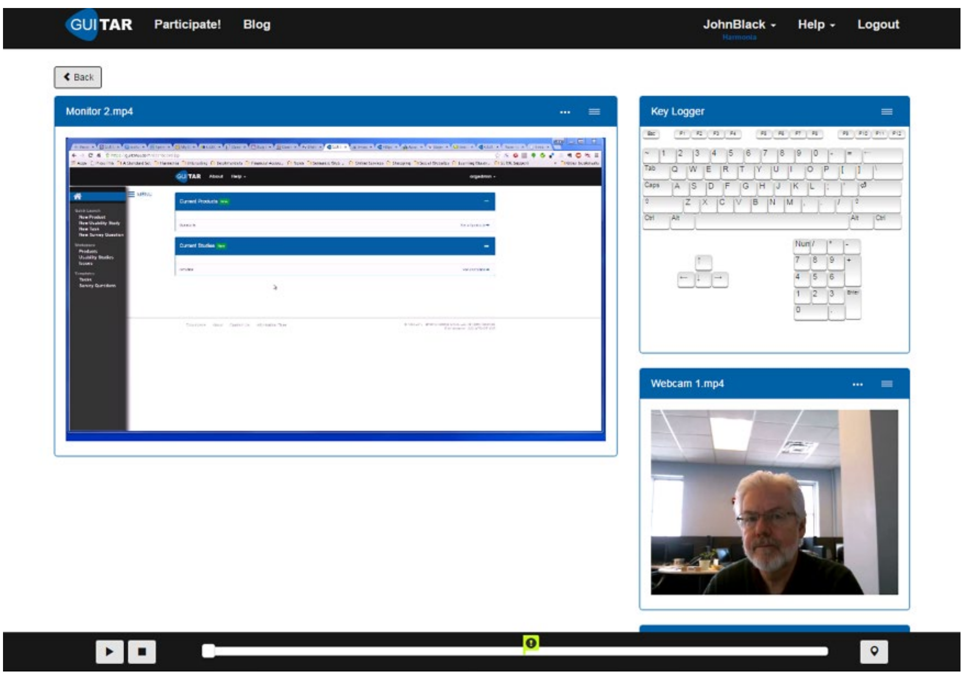


Figure 15.6 All recordings are synchronized when played in media player.

surveys such as the System Usability Survey (SUS) (Brooke, 1996). In addition, we include the capability to attach a survey question following each task in the test plan.

Survey questions suffer from certain limitations, which are described elsewhere, and the value and the problems with using them are not much different when performed remotely via another software mechanism.

Analysis and reporting The GUITAR Observer component is an example CAQDAS. The field of CAQDAS got its start in 1989 ("What is the CAQDAS Networking Project?" n.d.). It can be used as an approach to utilizing qualitative remote user research

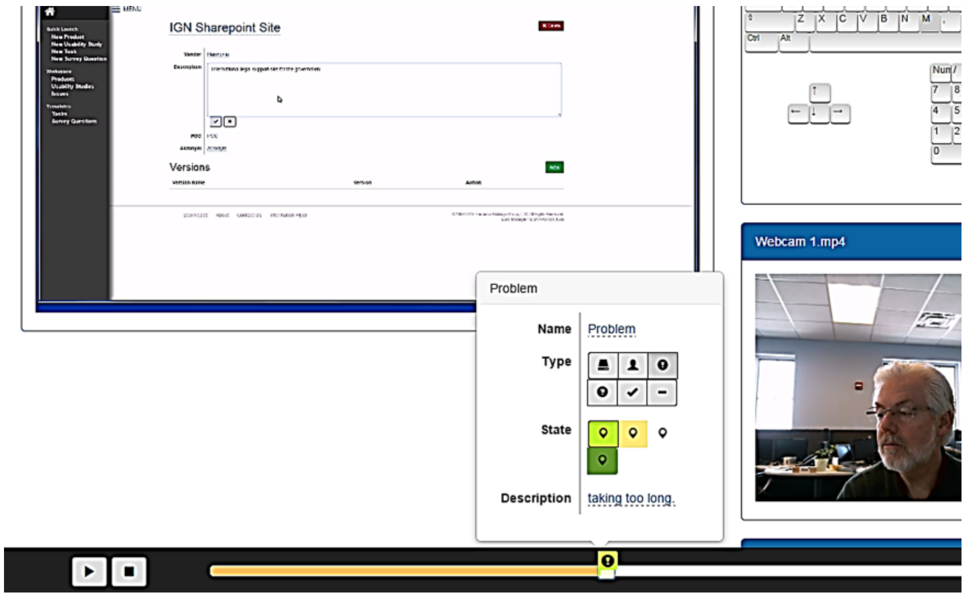


Figure 15.7 Any issues are tagged and saved with a time stamp.

data collected while doing remote usability testing. There are many products available that can be used, some are free, and some offer reduced cost to academic researchers.

Issues can be marked, as seen in Figure 15.7, at a specific point in time, with an identifier, assigned to a category, marked with the current state of remediation, and further described for reporting and communication. These issues can be pushed to a software development lifecycle development tool for tracking by developers or included in reports to stakeholders needing to respond to the problems identified.

Once the analysis and coding has taken place, either by an individual or by a select team of reviewers, the codes can be further collated and analyzed quantitatively. It should be noted that many usability problems reveal themselves to some evaluators but not others.

Analysis of qualitative recorded data There is a unique set of problems when doing an analysis of qualitative recorded data. Recorded data can be voluminous. Finding the valuable insights from a massive set of recorded data is a challenge. Out of a 10 minute video there may be 10 seconds that contains the most valuable and usable data. You need to be able to scan, isolate, and extract the valuable pieces from large recorded data sets. But to facilitate peer review, large volumes of data must be preserved.

Audio reports with valuable sound bites must also be transcribed to written text. At Harmonia, we are investigating how this could be facilitated by automated voice to text transcription. These transcriptions could then be analyzed automatically to look for key words and phrases, reducing the work of the researcher. Without these aids, experts I have interviewed report spending 10 times the hours used to record the data in extracting and analyzing it.

Issue fixing, remediation, and repair With the goal of automating the process of evaluating an emerging or existing product for its usability, Harmonia was faced with the problem of using the qualitative data recorded remotely to make recommendations for the removal of problems. The requirement was for this to be something that a nonexpert could accomplish. This is something that is especially important when the testing is done formatively, during the creation and development of the system.

Once testing is done, problems need to be coded, described, and communicated back to the developer of the system. The communications need to include what changes are recommended to make the product more usable. Presumably, the developers were not aware of the issue or they would have prevented it in the first place.

With remote user testing, the qualitative data recorded that reveals the problem must be quoted or shown to the developers to make them aware of the issue and the remedy desired.

As issues are located, recorded, and fixed, this progress needs to be monitored and tracked by project managers. The GUITAR tool can be integrated with common software issue tracker software such as Bugzilla (https://www.bugzilla.org/). When an issue is detected, the critical values can be automatically pushed to the development issue tracker.

Remote Testing Methods

Methods of measurement

There are many methods of measurement that are used in remote unmoderated user testing. Three of the most commonly used methods are these. First, there are measurable consequences of the absence of usability in a system. These range from failure of the product to be used by the user to succeed at some goal, all the way to death of the user, as is the case for some failures of an interface in military applications. Second, the target test subjects can be asked directly about their experience of the usability of the product, and their responses can be analyzed. Finally, an expert can review recorded media of the test participant using the system and use various heuristics to evaluate the usability of the system.

Consequences, the market, and usage statistics

In the commercial world, the market is the final arbiter of usability. Unusable systems will not continue to sell. In the military, problems can be diagnosed by even worse consequences (Sharon & Krug, 2015).

In a more morbid example of usability data that is collected, the black box that records the last words and system measurements of an airplane before it crashes, killing everyone onboard, can reveal problems with the controls inside an aircraft, which have sometimes been to blame for the death of hundreds of people. In such cases, where every participant is dead, these are only the recordings made to help diagnose the problem with usability of the interface.

Software that collects and sends usage data to the company or organization that produces it is part of another common remote usability research workflow. Most software also produces log files of various sorts. These log files, or core dumps, constitute yet another form of remote usability data collection practice that has existed for almost as long as computers have existed.

Opinions expressed versus behavior observed

When doing user research of any kind using a speak-aloud protocol, the goal is not to hear the user's opinion on the product. The goal should be to learn what they are actually thinking while they use the product. This may differ significantly from what they say they are thinking and doing. So when a speak-aloud method is used, beware that it could be misleading.

End-user reviews

Real end-user responses provide the most realistic form of measurement of usability. In the end, nothing else matters. Product reviews on public retail web sites are a prime example. Even here, there is a problem with fraud.

Asking survey questions

One of the primary means of detecting usability problems is to ask test participants about their experience. This is typically done with a Likert-scale question. The SUS survey, which is included in GUITAR, is an example of a survey that simply asks the test participant if they found the system usable.

Observing behavior directly

The field of ethnography is dependent on thoughtful observation of the behavior of an individual when the researcher is in the direct presence of that individual. What's different in the remote case is that the researcher is interested in observing only what can be obtained through a mediating technology.

Behavior can also be observed by viewing video recordings of individuals and actions. This again places another gap between the researcher and the observed participant. No longer can the researcher respond in the present to the events that he or she observes. And this has numerous consequences for the kinds of data that can be collected.

Remote Testing Metrics

Types of metrics

There are many metrics that can be collected remotely that can reveal usability issues. Among the most commonly used are these:

- time to complete a task;
- accuracy of the result produced by a task;
- number of mistakes made on the way to completion of a task.

Especially in remote unmoderated user testing, the data that can be collected tends towards numeric metrics of this kind. Because of the barriers and gaps separating the researcher from the observed participant that are in effect with unmoderated, remote user testing, it is often the case that the most value is gained from more simple metrics that can be measured automatically.

Task-completion duration

The time it takes for the test subject to complete a task is a common measure in usability studies. In most goal-oriented activities, it is important to complete tasks sooner rather than later. If a task takes a long time it can indicate difficulty.

Accuracy / correctness of the result

Even if the test participant completes a task quickly, the result may fall below an acceptable threshold of accuracy or correctness. Establish a measure of the accuracy or correctness of the task result for comparison. If the accuracy is low over many instances it could indicate a usability problem with the interface. But there are many possible reasons for low accuracy and incorrectness. One of them is the usability of the system. It could also reveal a system interface that is difficult to learn.

Number of mistakes made

In addition to the relative accuracy of the final result, a valuable metric is the number of mistakes that were made to arrive at a completed result. This is often called the number of retries. If the number of retries needed is high, it could indicate a usability issue even if the final result was totally accurate.

Converting qualitative events to metrics

The terms used to describe qualitative problems should, if possible, be translated into quantitative measures. For example, if a qualitative behavioral event can be described as "menu choice is hard to find," a metric could be created to measure the time it takes to find the menu item. When success on a task such as this is defined by a number, it is often useful to create another metric that identifies an acceptable time limit over which the first metric is considered an issue. This threshold should be defined by empirical testing of average users on this or a similar system. But this may not also always be possible at reasonable cost. In that case, it may be necessary to rely primarily on heuristics.

Goal fulfillment

Goal fulfillment is a metric that quantifies the degree to which the user's goal was achieved. Users are generally acting to accomplish a goal using the system interface and they either do or do not achieve it. Goals are a good way to structure and quantify higher level behavior, consisting of multiple tasks.

Other metrics

There are many other interesting metrics that you can collect remotely. For example, you can collect the most common path taken to accomplish the task, and measure how many steps it contains. You can collect the number of times the back button is pressed. You can record the total number of page views that are made attempting to fulfill a certain task or goal. You can track the dropoff rate for a task—that is, how many people abandoned the task and at what point.

Remote Testing Practice

Introduction to the practice of remote user testing

In this section I discuss some of the practical issues involved with remote usability testing. I consider recruiting test participants, providing incentives, validating that test participants are who they say they are, and that they really belong to the desired demographic group. I discuss collecting data for verification versus privacy considerations. I consider some of the problems faced in defining a fixed test plan for remote use, as well as how to solicit and ensure speak aloud reports. I discuss the ethical issues with recording screen displays and facial expressions in a test participant's native environment.

Using sampling

It is almost never possible, in formal research, to achieve the ideal of testing all actual users using a live system. Thus, just as in other kinds of quality assurance, researchers often attempt to find subjects that are as close as they can find to actual users, using the method of *sampling*. This requires identifying the attributes of actual users that are critical to the questions about the use of the product under investigation. For example, they may look for people that have a certain skill set, or certain language abilities, or certain reasoning maturity. There are many characteristics of the people they are investigating that can affect the usability they experience with a given product. Deciding which of those are critical to the investigation is the first step in conducting a study on a valid sample of the population of interest.

Understanding users and test participants

In order to use sampling, your test participants need to be as close to your target users as possible. But in order to evaluate how close test participants are to actual users, you need to know as much as you can about your actual users. In some cases this may be difficult as you don't know who exactly will adopt your product or even if it will be adopted at all, but, in the absence of data about actual users, there must be some assumptions and they must be used to establish a baseline for the characteristics that the test participants need to show. If actual users already exist, data about them should be collected to use to establish the typical characteristics needed for test participants.

Recruiting test participants

Recruiting test participants for remote usability testing is both easier and less reliable than it is for laboratory research. It is easier because the potential users can be anywhere in the world and can participate at any time of their own choosing. However, it is less reliable because it is harder to tell if the people who respond to your solicitations are really who they say they are. So the reach is inversely proportional to the reliability.

Businesses exist that provide a service for locating groups of people who meet specified demographic characteristics. Otherwise, you will be forced to use standard advertising methods to attract the test participants from the groups you want to study. To help with this, the GUITAR team is investigating partnering with companies that specialize in the identification, demographic sorting, contracting, and paying people in an on-demand business. One such company is Ethnio (https://ethn.io/).

Currently, the GUITAR product does not have the ability to facilitate recruitment. The Army customer felt that issues of privacy were far too complex and difficult to make it worthwhile.

You may find that it is necessary to use incentives to attract a proper group of individuals for your tests. In cases like this you may offer the value of the resulting product as an incentive. For example, if the application would help doctors be twice as productive they might find a free license to be a strong incentive to participate in the test. Otherwise, you may have to pay significant amounts to get the test participants you need.

Validating the demographics of participants

Once you have located a pool of users to act as test participants you will want to validate that they match the demographic criteria that you need. But verification of the demographics of your test participants brings with it additional problems of privacy that must be addressed. The more data you've gathered to validate who your test participants are, the more data you will have about them. This must be kept private.

No matter how you attract your target subjects the validity of your test could depend on your ability to vet your participants. Financial businesses, as well as others, use various techniques to validate that online, remote applicants are as they are representing themselves to be.

Verification of data versus privacy

There is an inversely proportional relationship between the accuracy of the demographic data that you collect on your test participants and the privacy that you need to provide them. The validity of your tasks may require the collection of personal data, which makes privacy of the utmost importance. If you were testing the usability of an application that helped people who were suffering from an addiction or other medical issue, for example, you would be required to keep that data private. But the same can be said for income levels or sexual preferences or any other kind of data that is especially sensitive. You should not collect the data until you have dealt with the security issues related to keeping such data private.

Maintaining control of the test

The main tool you have to ensure that your test-plan instructions are clear and can be followed is through testing. You should test your test plan with representative users before you release it to the public. This cannot be overstated. Don't release a test plan without running tests to make sure it is clear and complete.

Eliciting and using speak-aloud reports

One of the basic techniques of user testing is to instruct test participants to speak out loud while they carry out the tasks in the test plan. And one of the problems faced when doing remote unmoderated user testing is the inability to remind the participants to speak out loud if they fall silent. It would be helpful when doing remote unmoderated testing if the software could detect silence and either stop recording or provide an automated prompt to get them back on track. For the GUITAR tool, we have looked at adding such a facility.

Another problem that arises involves distinguishing between conversations that are not related to the task execution but may be recorded inadvertently. For example, if the test participant received a mobile phone call during task execution, there may be lots of verbal expressions that have absolutely nothing to do with the task. The ability to distinguish such extraneous data from valid and important speak-aloud reports is one of the challenges of any type of remote user test.

Deleting recordings on request

Facial images, screen displays, audio tracks, or keystrokes may all contain personally identifiable information. If recorded in the participant's native environment, that information could be something the test participant wants to keep private. Participants should therefore be able to have all their data deleted should they realize that an event took place that they do not want published.

Age limits and restrictions on minors

Along with privacy and security measures in general, you have must take special precautions when dealing with minors. The Children's Online Privacy Protection Act of 1998 (COPPA) is a federal law restricting what information minors under the age of 13 can provide without parental consent.

Storing user test data

Once data has been collected it needs to be kept private and secure for as long as it exists, so determining the valid and necessary lifetime of any data that is collected should be dealt with up front. Massive amounts of data could be collected. Storing it can be expensive. Planning should be done before running tests.

Other issues related to data must also be planned for as well. For example, who will own the data when the study is done? Does the company that provided the user testing service own the data or does the company that hired them? And does the individual test participant have any right to recordings in which they appear? If the company that is storing the data is sold or closes then what is the disposition of the data? All of these questions should be considered and plans made to account for them.

The Future of Remote Testing

A vision of the future of remote user testing

Ultimately, the best user research would be to test all actual users, during actual use, in an ordinary context, in real time, without any disruption of the user due to testing. This is not currently possible but, remarkably, researchers seem to be quickly moving in the direction of this goal—using the Internet, biometric user identification, wearable sensors, ubiquitous recording devices, and wireless power and data transfer.

Put together, researchers are not only going to be able to conduct summative studies on finished products, but when incorporated into the software development lifecycle, using continuous integration, deployment, delivery, and testing, they could achieve something like customized, personalized systems interfaces delivered on demand, which have proven usability for each individual and context built in from the beginning.

Key Takeaways

Summary

The worst problems are very easy to find. Remote testing can be used to find them quickly. Every user and situation in which a system is used is different and can affect usability. Continuous real time testing of all users is the ultimate goal. Systems that automatically conform to each user and make them comfortable are coming.

The worst usability problems can be discovered easily

Even simple, unmoderated remote usability testing can uncover the worst usability problems that affect products. You can complete it fairly quickly with a few test participants. Many of the worst usability problems with the system are so glaringly obvious that they are nearly instantaneously identified with the heuristic evaluation of collected recordings of an end user. For this reason, all products should be tested for usability.

Every instance of human system interaction is unique

Every user is unique. What is usable to one may not be usable to anybody else. This is because every instance of human interaction with a system is unique. Any user will not be the same on any two different occasions. Their cognitive abilities are constantly shifting as are their attitudes and beliefs. And conditions in the environment may be shifting as well. For this reason, it may not be possible to be able to say once and for all this system's interface is usable. In many cases, the best a researcher can do is to make a prediction. And that prediction should always be couched in terms of a specific type of user and specific context of use.

Use all techniques—including the lab

The goals of usability and user testing are to produce better human-computer interfaces. And the increases in quality that can be gained from testing are great. But there many different methods that can be used to collect large amounts of data related to humans use of computers. Because of this, user testing should include many of these different techniques, based on their suitability to the actual context of use.

GUITAR from Harmonia is designed to act as a central hub and repository of as many quantitative measurements and qualitative data sets as possible. Many researchers in the human systems integration space learn to use one research technique, and then they keep using it even when there are other techniques that may serve them better in other circumstances. We are working to make GUITAR capable of evaluating test plans to determine whether research techniques that are *not* included might help produce better results. By analysis of the goals desired, recommendations can be made to include other methods in the research plan.

References

Anderson, R. (2011, April 1). User (experience) research, design research, usability research, market research … A changing, interconnected world. *UX Magazine,* article no. 648. Retrieved from https://uxmag.com/articles/user-experience-research-design-research-usability-research-market-research

Bolt, N., & Tulathimutte, T. (2010). *Remote research: Real users, real time, real research* (1st ed.). Brooklyn, NY: Rosenfeld Media, LLC.

Brooke, J. (1996). SUS: a "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & I. L. McClelland (Eds.), *Usability evaluation in industry* (pp. 189–194). London: Taylor & Francis.

Nielsen Norman Group. (2014, June 1). *Selecting an online tool for unmoderated remote user testing.* Retrieved from https://www.nngroup.com/articles/unmoderated-user-testing-tools/

Rochanayon, A. (2015, August 14). Top 7 remote testing methods you're almost definitely not doing [Web log post]. Retrieved from http://www.userzoom.com/remote-unmoderated-usability-testing/top-7-remote-testing-methods-youre-almost-definitely-not-doing/

Sharon, T., & Krug, S. (2015, May 19). Remotely Possible Episode 1—Remote unmoderated usability testing [Video file]. Retrieved from https://www.youtube.com/watch?v=4yZWrHsmNtY&index=1&list=PLxjM5m7-ZGWJ3ltBYig0cA-ii630wRUOF

Varga, E. (2011, October 20). *An experiential comparative analysis of two remote usability testing methods* (master's thesis). Rochester Institute of Technology, Rochester, NY. Retrieved from http://scholarworks.rit.edu/theses/691

What is the CAQDAS Networking Project? (n.d.). Retrieved from http://www.surrey.ac.uk/sociology/research/researchcentres/caqdas/

16

Applied User Research in Games

Randy J. Pagulayan, Daniel V. Gunn, Jerome R. Hagen, Deborah J. O. Hendersen, Todd A. Kelley, Bruce C. Phillips, J. J. Guajardo, and Tim A. Nichols

Introduction

Applied games user research has a history dating back over 40 years of adapting social science methods (i.e., psychological, sociological, anthropological) to the game development process with the goal of evaluating and improving player experience (Amaya et al., 2008; Medlock, 2014). Since the late 1990s, User Research groups at Microsoft have been applying, refining, and inventing new research methods to improve both the usability and "enjoyability" of our game titles, with many other game development studios and publishers to follow (e.g., Sony, Activision, EA, UbiSoft). The purpose of this chapter is to share some of the philosophies we have developed over that time and expose best practices for applying research methodologies in game development across key topic areas. First, we describe the history of the discipline and how we define games user research. Next, we describe our approach to games user research and how it fits into the game development cycle. Finally, the chapter will conclude with concrete examples in a few focus areas of the discipline and lessons learned along the way.

History

Beginnings (1970s)

Games user research can trace its origins back to the early 1970s and the work of the very first games user researcher, Carol Kantor (see Medlock, 2014 for a detailed historical record). Carol began working at Atari in 1972 and was tasked with trying to figure out a way to determine if a particular game in development was going to be a success or failure. Much of her early testing was on Atari and competitor pinball games but with the burgeoning arcade video game market of the 1970s, she quickly moved on to testing video games. Carol leveraged her marketing research background

The Wiley Handbook of Human Computer Interaction, Volume 1, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

and adapted and developed several user research methods that are still in use today. There were four key methods Carol used:

- Field observation methods—coding of player behavior in arcades.
- Survey methods—approaching people playing a game in the arcade and having them fill out a questionnaire about their experience.
- Focus group methods—bringing in members of the target audience during game development and having them play the games and participate in a focus group.
- Coin collection data—analyzing the number of quarters accrued per title.

For several years, Carol did much of the testing herself. In 1977 and 1978, she hired Colette Weil, Mary Fujihara, and Linda Benzler. Together, these four constituted the first team of games user researchers and their work continued on into the 1980s.

1980 onward

Following a paucity of growth in the discipline in the 1980s, the 1990s were a watershed decade for games user research. Major companies began getting involved in it and several companies were founded to support it. In the early 2000s, the games user research industry really began to pick up momentum. Several teams were established at major game developers, researchers in the industry began to publish, and numerous efforts to mature the discipline could be seen. In our current decade, we continue to see the establishment of games user research groups at major companies. In addition, we also see increasing community outreach as well as momentum building on the applied academic front. See Table 16.1 for notable moments in games user research.

Looking back, to look forward

Games user research has grown tremendously since Carol Cantor and her team began conducting research on pinball machines and early video games over 40 years ago. As we look back over the past 20 years of the industry, we have a sense of awe and pride in what the industry has become as well as a healthy appreciation for lessons learned from our own experiences. For the remainder of this chapter, we will describe how we define games user research and how to apply methodologies in the games development cycle. We will also describe our progress in some specific areas of the field that are of critical importance to the success of a great gaming experience. The specific areas we will discuss are:

- testing narrative in games;
- providing design intelligence;
- conducting user research with kids;
- user-generated content;
- user research in motion gaming;
- the lure of biometrics.

It is our hope to provide inspiration to researchers to leverage, build upon, and improve existing techniques to help carry games user research into the future.

Table 16.1 The history of games user research.

Decade	*Year*	*Notable moments in games user research*
1970s	1972	First games user researcher, Carol Kantor, hired by Atari
	1978	Carol Kantor, Colette Weil, Mary Fujihara, and Linda Benzler become the first games user research team
1990s	1992	Brenda Laurel begins ethnography research on girls and gaming while working at Interval Research Inc.; Nicole Lazzaro founds games consulting company XEODesign with a focus on games user research; Heather DeSurvire founds Behavioristics, a games user research consultancy.
	1994	Amy Kanerva and Libby Hanna run the first traditional usability studies on a Microsoft game—Magic School Bus.
	1996	Sony Japan (SCEJ) creates a game "monitoring" team run by Shuhei Yoshida. Mark Cerny begins a similar practice with Sony America (SCEA) where he conducts difficulty balancing tests with participants; Brenda Laurel founds game development company, Purple Moon, and uses field research and other techniques to develop games targeted at young women.
	1997	Bill Fulton is the first dedicated user researcher working on games in the Microsoft Games usability group; Howard Phillips is hired at Microsoft Games to be the playtest manager and asks Bill to run playtesting efforts; Michael Medlock replaces Bill in the games usability group; Bill moves to the Microsoft playtesting group.
2000s	2000	The Microsoft games usability team combines with the Microsoft games playtest team and becomes Microsoft Games User Research.
	2002	First games user research presentation at GDC, *Beyond Psychological Theory: Getting Data that Improves Games*, is presented by Keith Steury and Bill Fulton; Melissa Federoff publishes the first set of game usability heuristics in *Usability Guidelines for the Creation and Evaluation of Fun in Video Games*; Tracey Seller founds the User Experience Lab—a games usability consultancy based in New Zealand; Sony America (SCEA) builds dedicated user research labs.
	2003	User research groups are founded at Ubisoft Paris (Lionel Raynaud) and Montreal (Stephane Cardin).
	2004	Janus Sørensen starts an IO interactive games user research team (part of Edios & Crystal Dynamics).
	2007	First time mainstream tech press acknowledges the role of User Research in games, *Halo 3: How Microsoft Labs Invented a New Science of Play*, is the cover story in Wired Magazine; Ray Kowalewski starts a user research group at Activision.
2010s	2010	Brooke White starts a games user research team at Disney Interactive; Jason Yow starts a games user research team at Disney Core; 1st IGDA games user research Summit is held with 40 attendees.
	2011	Microsoft Games User Research becomes Microsoft Studios User Research with an increased emphasis on analytics and research operations; second IGDA games user research summit is held with 76 attendees.
	2012	Dennis Wixon becomes the first professor of games user research at USC; third IGDA games user research summit is held with 97 attendees.
	2013	Fourth IGDA games user research summit is held with 130 attendees.
	2014	Fifth IGDA games user research summit is held with 150 attendees.
	2015	Sixth IGDA games user research summit is held with 185 attendees; first annual European games user research summit is held with 93 attendees.
	2016	Seventh IGDA games user research summit is held with 193 attendees.

Sources: 1. Medlock, M. C. (2014, September). *History of video games user research*. Paper presented at the 2014 Games User Research Summit, San Francisco, CA.; 2. Personal communication A. Abney and D. Tisserand, August 1, 2016.

What is Games User Research?

Applied versus basic research

At the most rudimentary level, research can be classified into two broad categories: basic and applied (Stanovich, 2007). Although not a truly dichotomous distinction, it does help to think of these categories when describing research in games because the approach a researcher takes will be guided by the end goal. Generally, basic research is conducted with the end goal of improving our understanding of a phenomenon—whether that be to get more insight into how something works, to help build a theory, to test a theory, or to extend a theory, and so forth. In the games space, examples can be seen in academic research focusing on how and why people play games (Olson, 2010). Improving a specific consumer product by applying the fruits of academic research can be done, but it is often not the primary focus of those researchers. Indeed, it can be challenging to directly apply the insight gained from basic games research into products (Hopson, 2006). In applied research, on the other hand, the primary goal is to apply the findings to improve *something in the real world*. For example, much of the human factors research in the aviation and automobile industry focuses on improving human interaction, safety, and performance with those products (Wickens & Hollands, 2000). In applied games research, the researcher typically tries to answer a specific set of research questions and the answers will have a direct impact on the design of a game (Amaya et al., 2008). The goal is to turn insight generated by research into improvements in our products—our products being entertainment experiences. Towards that end, our processes, infrastructure, and philosophy are designed around helping development teams make the best user experience possible for their products by leveraging games user research.

The three keys to games user research

We define games user research as the application of controlled research methods to the development of computer and video games to ensure an engaging, approachable, and fun player experience. Although the definition is straightforward and simple, the actual process is far from it. Conducting research and applying the findings to improve a game within the confines of the fast-paced game development environment is no small undertaking. In contrast to some nonindustry research where the data and interpretation can be done over the course of years or months, in the applied games user research, the timeframe is more likely weeks or days. From our experience, there are three key ingredients that are required to make the process a success. If any one of them is compromised, the ability to use the findings of our research to improve the player experience, will likewise be compromised. Those key ingredients are firmly seated in the term *research*, so it is important to break that term down even further.

Leedy and Ormrod (2004) describe *research* as the systematic process of collecting and analyzing information to increase our understanding of the phenomenon under study. According to the authors, it is the function of the researcher to contribute to the understanding of the phenomenon, and to communicate that understanding to others. Looking at this definition, three important key aspects can be pulled out: the first key aspect is the RESEARCHER, whose function is to contribute and communicate to others; the second key aspect is about the METHOD, a systematic process of collecting and analyzing information; finally, the third key is what we call the OUTCOME, which is to increase our understanding of the phenomenon, but for the purposes of positive change. For us, games user research is about the intersection of those three key elements: researcher, methods, and outcomes (see Figure 16.1).

For games user research to be successful, all three elements need to be sound. Looking at various combinations of the keys, it is evident how some mixtures will inhibit the ultimate goal of applying research results to improve the player experience of a product (see Figure 16.2). For example, an unqualified researcher leads to poor methods, which produces misleading or damaging results (see Figure 16.2a). Or maybe you have someone who is a specialist, who can do great research in one area, but they get pushed or try to do methods beyond their expertise or ability (Figure 16.2b). Finally, if you have a good researcher who executes on the methods correctly, but they can't communicate back to the development team in a way that people understand, the results can't be successfully applied to the betterment of the product (Figure 16.2c).

The researcher What makes a great games researcher? From our experience, great games researchers must have a core skillset to be effective (see Figure 16.3). They must have the ability to take things that are abstract and difficult to measure, and make them concrete and measurable. Measuring a player's perception of fun, pace, or challenge is not a trivial task. Yet, the researcher must be able to quantify, and measure, these constructs in addition to any other parts of the player experience the designer is trying to create (Pagulayan et al., 2012). Researchers also must be able to take speculation and turn it into testable hypotheses. Designers often rely on speculation to make design decisions when no other data are available to leverage. Great researchers can take a designer's speculations about how players might react, or what their experience might be like, and turn that it into testable concrete scenarios.

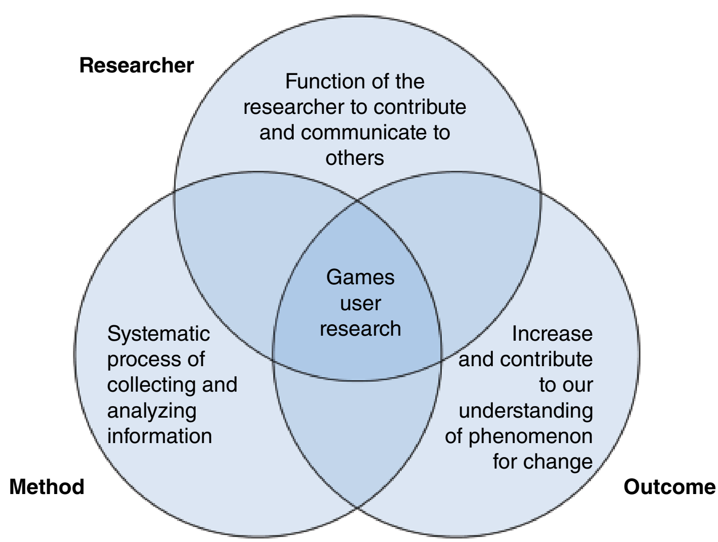


Figure 16.1 Three keys to games user research.

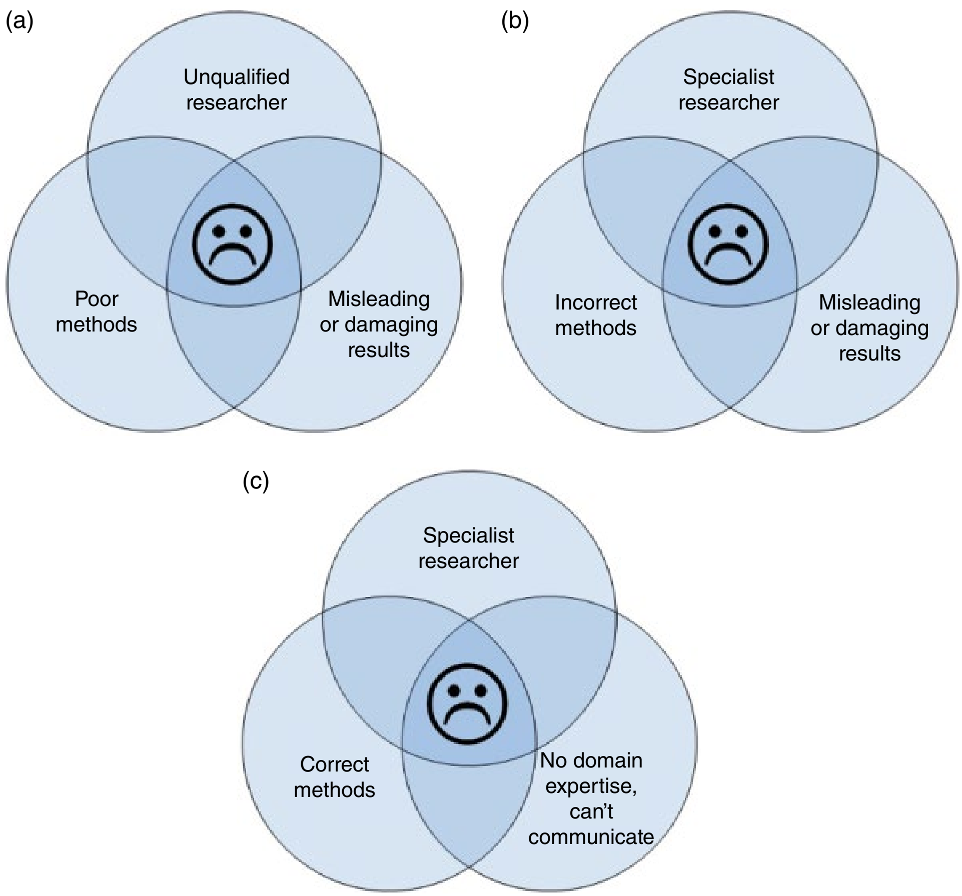


Figure 16.2 Some key combinations that lead to suboptimal results.

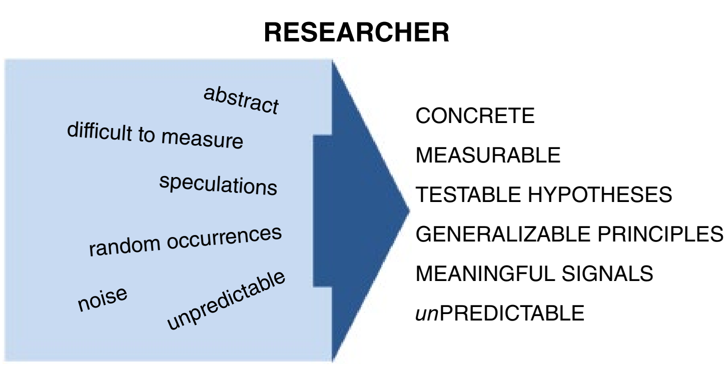


Figure 16.3 Skillset required for researcher in games.

Researchers must also be able to take random occurrences and make them generalized principles. For example, to an unseasoned researcher, player behavior in response to a specific product might seem tied just to the nuances of that product experience. However, good researchers can see patterns across testing of a variety of products and pull together the underlying principles (Herbst & Medlock, 2014). Good researchers must also be able to see the signal within the noise. Some techniques for measuring human experience and behavior can produce a tremendous amount of data (e.g., game telemetry, usability, biometric techniques), and it requires a researcher who can identify the signal correctly and feed it back to the design team. Lastly, great games researchers must be able to take the unpredictable and make it predictable. With experience, a games researcher begins to be able to predict the outcome of certain design decisions before any testing begins. Of course, researchers cannot predict all outcomes (otherwise there would be no reason for testing) but, through experience with testing users across numerous games and a background in the behavioral sciences, a great games researcher can apply certain heuristics to evaluate designs and tell the team the likely outcomes of testing before it takes place.

In our experience, the most successful foundation for being a great games researcher is training in experimental psychology. Most of the applied techniques we see in industry have their origins in psychological research methods (e.g., usability testing) and if one has training in experimental psychology, they can easily pick up the applied methods. Sometimes even more important is the researcher's ability to adapt new methods to the unique challenges we see in entertainment (e.g., Motion-based gaming, Narrative); this is harder to do without a foundation in experimental psychology.

Method What makes a great method? There is no single great method for everything in games user research just as there is no single great tool for every woodworking project. For example, a hammer is great for hammering nails but not so great for cutting through plywood; a tool's goodness varies with the particular project at hand (as well as the tool's operator). One factor in determining a method's goodness is its ability to answer research questions. A good method is one that allows a researcher to answer the design questions in a way that leads to actionable insights. There are a variety of methods used in games user research (Amaya et al., 2008; Pagulayan et al., 2003, 2006, 2012) and some of these were developed in games (e.g., RITE testing, playtesting) and others were not (e.g., contextual inquiry, focus group testing). Typically, the researcher will select the best method to answer the question at hand—with the research question driving the method chosen. It is important that the process should not be reversed (method driving what research question to ask). The research question must drive the design of the test and methods used, which drive the data, which drives the analysis, which drives the interpretation, which eventually drives improvements to the game and to the user experience. In situations where methods don't exist in games research to answer the research questions at hand, we pull from our experimental psychology foundation to invent new ones. As in the tools analogy, we also categorize our methods according to how they can be used, but instead of the categories being "woodworking" or "striking" or "cutting" tools, our language revolves around our methods being attitudinal, behavioral, physiological as well as quantitative or qualitative. For a more thorough discussion of our methods, see Amaya et al. (2008).

Outcome The final key to successful games user research is what we call "outcome," or effective communication of results for the purposes of change. Good research is only effective if it is understandable and actionable. This requires two things: domain expertise (being a gamer) and the ability to communicate. The person conducting the research should understand games and entertainment from personal experience; this is not a space where research methods can be effective without understanding the nuances of the experience. In addition to understanding the domain, the researcher must be able to communicate effectively. We have experienced situations in which a solid researcher, conducting solid research, does not successfully help improve the game because of failed communication. Communication must clearly convey the results and implications to the game development team research audience. Research can be full of caveats and designers need to be able to get the answer to the question they are looking for before they can apply the insight to the design process. The analysis must help them make informed designs decisions and not make the problem more complex. The most effective researchers can communicate and influence in a way to make the output consumable and usable.

Our Approach to User Research in Games

Accurate user feedback loop

Imagine you are a game designer working on a game. What would be some of your overall goals for that game? If you listed out your goals, you'd likely have several that either implicitly or explicitly referenced a player—the game is fun, the game is challenging but not impossible, the game feels like a mix between Gran Turismo (1997) and Pokémon (1996), and so on. If you were then asked a clarifying question such as, "The game is fun for whom?" you would likely say, "the target audience." Most designers hope that their creation will be fun for the target audience. For example, if they are building a racing game, they want it to be fun for racing gamers, if they are building a first-person shooter game, they want it to be fun for first-person shooter gamers. However, throughout the development of their game, it becomes difficult for designers to assess how close they are to achieving their goals without an accurate feedback loop from the target audience.

As Fulton (2002) points out, it is not that designers have a lack of feedback about their game to act on, it's that they have too much and don't know which channels to listen to. Within the games industry, game designers are constantly receiving feedback about their games in development from multiple sources: friends, family, other designers, producers, artists, programmers, neighbors, and so on. There are two major factors that need to be considered when looking at how to interpret that feedback: (a) how representative the feedback is of the target audience; and (b) how the feedback is gathered. For example, the designer's best friend has likely seen many more in-development racing games than the typical racing gamer, and they are the designer's best friend. When asked what they think of a new track that the designer spent the last month designing, their feedback will likely be biased.

At a very general level, the role of a games researcher should be to provide an *unbiased* feedback loop to game designers. Just as anybody working on a project for long periods of time loses perspective, designers also see their game differently from a majority of outsiders and, more importantly, different from the typical player. It is our job to ensure that development teams get a perspective on their game through the lens of a typical player. Towards that end, we collect unbiased data from real users and use that data to improve our games in development. However, the process is not as simple as bringing in the target consumers and just asking for their feedback. We must leverage our expertise in psychological research methodologies to mitigate as much bias as we can. For example, the setup of the testing, the interactions players have with the researcher, and the interactions between players all have to be carefully controlled to ensure data integrity. Again, the reason to focus on hiring researchers with a background in experimental psychology is the training to gather data from people in a way that minimizes the potential for bias to creep in.

When providing an unbiased feedback loop to the development teams, it is also our job to ensure that other feedback loops they are leveraging are put into proper perspective. Teams read social media and online forums for feedback about their games but those feedback loops often suffer from several well-known threats to validity (e.g., self-selection and nonindependence of data). We have demonstrated to teams that, at times, the strong messages they are getting from online forums suggesting design changes actually don't represent the vast majority of the audience's sentiment. In those cases, acting on that feedback could be detrimental to the game overall. Thus, it's our role to give voice to the quieter majority they don't hear from on forums and ensure teams put these other feedback loops into proper perspective. To do so, we must first establish a relationship of trust with our development teams.

Importance of the researcher / game developer relationship

In the world of applied research, research must be both high quality and effective. This requires a separate enough relationship with the game development team for the research to be objective but a close enough relationship with the game development team to understand context and for the research to be listened to.

In practice, we refer to this balance as embedded but separate. Researchers should be an embedded part of game development teams throughout the game-development process. This allows researchers to understand deeply the team's design goals and the player experience they're trying to create. This embedded relationship means that researchers can affect the direction of game development from the start and build time for research iterations in the schedule, as many issues can only be addressed by integrating research insights into the entire game development cycle. And this close relationship also helps researchers build trust with game development teams.

It's also very important that researchers are separate and objective, which we achieve through our location in the game development organization. Being in a centralized organization that supports all of the game development teams means that we don't answer directly to any game team or game studio, and our top priority is always to be an advocate for the user and the user experience. The organization as a whole relies on researchers to be unbiased sources of data and research insights who can dispassionately tell them how their games' user experiences compare to player expectations, design vision, competitors, and the organization's game portfolio as a whole.

A centralized location in the organization also gives researchers a vantage point to look across all products and draw research insights and comparisons from all of them. This centralized organization separate from game development teams is extremely valuable to help researchers resist the groupthink that often happens when all the game development team members are completely immersed in game development on a single title.

Application in the game development cycle

Game development has a finite duration and finite available resources. No matter how much everyone would like to, there isn't the time or the money to address all the player experience issues that are discovered in research, especially if they're discovered late. Different development activities occur at different stages of development, and early decisions often get locked into place because of the production needs of the game development process. For instance, the physical space of gameplay levels often needs to be constructed early so that artists can craft the appearance of the space, gameplay designers can design the gameplay that goes into them, and other development team members can add their contributions.

As researchers, this means that it is critical to understand the game development process deeply and focus on the prerequisites for a great user experience at the time when they'll have the greatest value. This also means conducting research early enough so that any issues can be addressed with time for additional research iterations to verify that the fixes had the desired effect.

There are five main phases in a typical game development cycle, each covering the creation of different parts of the player experience:

1. **Concept:** The focus here is on creating the basic idea for the gameplay that will make players love the experience and keep them wanting to play it. Typically, the concept phase only involves documents and images, so no interactive gameplay has been built yet.
2. **Prototype:** This is all about creating a gameplay prototype that demonstrates what's going to be so enjoyable about the experience. The focus for a prototype is often on unproven or otherwise risky gameplay experiences, because those experiences are critical to player enjoyment, and they usually take the most iteration to make them work well for players.
3. **Preproduction:** Here the focus is on preparing for, and designing, the player experience for the full game. From an experience perspective, this usually means iterating and expanding on the experience that was created in the prototype, and figuring out what will be necessary to create a successful complete player experience.
4. **Production:** Production is about executing the plan for building the whole game so that it's ready for release. Beta releases often happen in production, which means that in-development versions of the game are released to members of the public. But with or without a beta release, this is a huge time for iterating on the game experience, as this is when most of the game is built.
5. **Postrelease:** This is after official release of the game to the public, but player experience iteration doesn't necessarily stop here. Many games continue to add and adjust gameplay content after release, and it's crucial that both new and existing players continue to love the game through those gameplay experiences.

User research focus in the development cycle

The default development process in the gaming industry has often led to major player experience problems, a lack of time to address these problems once they're identified, and a lack of flexibility because previous decisions limit options for how to address the issues. It's become clear that effective research is not just about choosing the right time to fit research into the development cycle. It's also about helping the development team plan out when and how to focus on building specific parts of the experience to produce the best chance of having a successful player experience.

As a result, we've created a sequence of five user research focus areas that are organized around what we've learned to be the critical components of a successful player experience (Hagen, 2015). These focus areas are organized to get the most out of research at each phase of development, and they help game development teams anticipate and remove roadblocks to player experience improvement as early as possible so there is enough time to address them. Each focus area is dependent on the previous ones to create a successful player experience. They are all dependent on the game development team accounting for them in the development schedule.

Audience definition (concept to prototype phases) Audience definition is about formalizing a shared definition and understanding of the target audience. Each game is different, and creating a successful experience is as much about who the game is being built for as it is what is being built. Different audiences have different preferences, desires, needs, and motivations, and they often enjoy different kinds of gameplay. Audience definition is a two-way feedback cycle between the game development team and researchers. One part involves the game development team defining the intended audience for whom the game is being made. The other part of the cycle is researchers feeding back data and research insights on the target audience to ensure that the fundamental design for the game is a good match for the players it's being made for.

Core mechanics (prototype to preproduction phases) One prerequisite for having a great player experience is that players should be able to understand and perform basic actions related to the gameplay experience successfully. This is about creating a solid foundation for gameplay by focusing on player success with those core gameplay actions, systems, and choices that will keep them having fun and wanting to continue playing the game. As mentioned above in the Prototype phase, it is important to understand whether unproven or otherwise risky gameplay experiences will work for players as early as possible. This is the first point in the development process when players are interacting directly with the gameplay experience, which leads to discoveries and improvements based on how the target audience actually interacts with the game.

Learning systems (preproduction to production phases) A gameplay experience is not a great one unless players actually discover and learn the components necessary to having fun with the game. Learning systems cover anything that helps players learn how to have that great experience, such as learning controls, core mechanics, gameplay systems, goals, or strategies for success. The learning systems themselves can range from subtle cues and gameplay situations that help players learn on their own to more explicit player teaching, feedback, and reminders. In a typical game industry development schedule, learning systems work often comes too late to allow for the iteration needed so that players learn successfully and have an enjoyable time while doing so. It is not an easy task to successfully create a fun system that helps all of the game's players learn and remember everything they need to throughout the game experience.

Initial experience (production phase) Once it is established that players can successfully play and learn the game, it is necessary to verify that they actually enjoy their first time picking up and playing the game, and that the initial enjoyment is enough to lead them to choose this game over all the other game choices they have. The initial experience of a game is when players decide whether the game is worth getting into. A bad experience can turn them off, or a great experience can draw them in. In contrast with how players will experience the game, game development teams often build the initial experience later than experiences that come several hours into gameplay. But even if the fifth-hour experience is great, players will never get to it if they didn't have a fantastic time in the first hour. That's what makes early building, research, and iteration on the initial experience so important.

Extended experience (production to postrelease phases) Once the initial experience is enjoyable and engaging, extended experience focuses on making sure the major game systems are compelling and balanced throughout the entire game experience. Additions and changes to the product also need to make players feel like they want to continue playing, whether those changes come before any public release, during a beta, or in postrelease. For some games, the extended experience will last for months or years of play. Extended experience is the focus area where our design intelligence efforts (see below) plays the biggest part as a data stream. Aggregated and easily analyzable player behavioral data combined with attitudinal data play a big part in balancing an entire game and making sure that it is enjoyable throughout.

The framework we have outlined so far describes our overall approach to testing: how we organize, when we focus on particular issues, and how we align to the needs and goals of the game design teams. However, during our 20 years of experience, we have dealt with many problems and challenges that reach beyond what is "standard" for games user research. The sections below describe some of the methodological and content testing approaches we've developed to address these challenges. The approaches include techniques to effectively do research on certain kinds of game experiences, do research on games for specific audiences, and to incorporate data streams based on what the data stream is best for.

Testing Narrative in Games

While traditional games user research has focused on gameplay and UI, game development teams frequently intend narrative to be a substantial portion of the experience. Designers want players to fall in love with the world, to love and mourn characters who are lost, to feel a sense of risk and challenge that motivates gameplay, and to come away feeling as though they have accomplished something grand. Unfortunately, game narratives don't always deliver on these goals for players.

Early attempts at testing narrative

When Studios User Research first started looking into game narrative, we brought a handful of participants into the lab and interviewed them about their favorite game narratives. One of the things we asked them to do was to retell their favorite game story. Here one participant tries to retell Skyrim:

> Umm, basically it started out I got my first dragon shout and then I went to the Greybeards and it taught me my first, eerr it actually taught me the first shout thing. And then I kept going umm what was the next part…ah it's so hard to remember the whole thing. Um what else I just kept. No there's this other dragon, ah, I can't remember their names…There's this other dragon and I killed him, and then there's one I didn't kill, and the one I didn't kill was, I can't remember… He was like, I don't even remember, honestly, it's like, it's really hard. (Participant 7)

This retelling was typical of what we saw from games. The players started off strong but struggled to make it from beginning to end (even on the games they selected as their personal favorite for narrative). By contrast, participants did not exhibit this issue when asked to retell their favorite story in any other medium (books, TV, movies, etc.). Here's that same participant retelling the story of *The Walking Dead* TV show:

> Um in the first episode he gets um he gets shot and then he's in the hospital and then he wakes up and he's just like whoa what's going on? And then he like sees Zombies and then he meets with his family after he goes to the town. I don't know what town it is. He meets his friends and then he meets up with um…god it's been a while since I've watched it. […] then the girl, she ran away at the end of the first season and they're still looking for her and that's kinda the whole first six episodes and then the second season they're looking for her plus uh the kid got shot. Then they have Walkers in the barn; they find out at the end of the second season which blew my mind. That was crazy! I was super freaked out about that. (Participant 7)

With each second retelling a pattern became apparent—while players stumbled to retell game narratives, this wasn't due to a lack of ability on the participant's part. Players were perfectly capable of remembering and retelling lengthy nongame narratives, from beginning to end. So if it wasn't the player, we had to conclude that something was going wrong with games (Hendersen, 2014a).

Finding those moments where a design goal has not been fully realized is an invitation for games user research to step in and engage. Diagnosing the gap between the game stimuli and the player's experience is precisely what we're best at, particularly when, as in the case of narrative, both the design expectations and financial investment are high. Narrative delivery devices like cut scenes are very expensive and if they are not landing the experience they are responsible for, we should help designers understand what's happening.

The problem

The challenge is that, while early forms of gameplay are available and playable as early as the prototype phase of development, narrative often isn't included in playable versions of the game early enough for meaningful research. The challenge was in figuring out what to research. While it is common to begin testing gameplay early on, that's because what's missing from gameplay doesn't overly impact the research focus areas. Enemy behavior is the same, even if the walls are the wrong color. For example, levels in Halo 4 (2012) looked very different in early and late research (see Figure 16.4):

The reason we do research with such early prototypes is that we can get lots of information about level layout, combat setups, weapon balance, and so on, and we can get this information in time for the team to make cost-effective changes. By

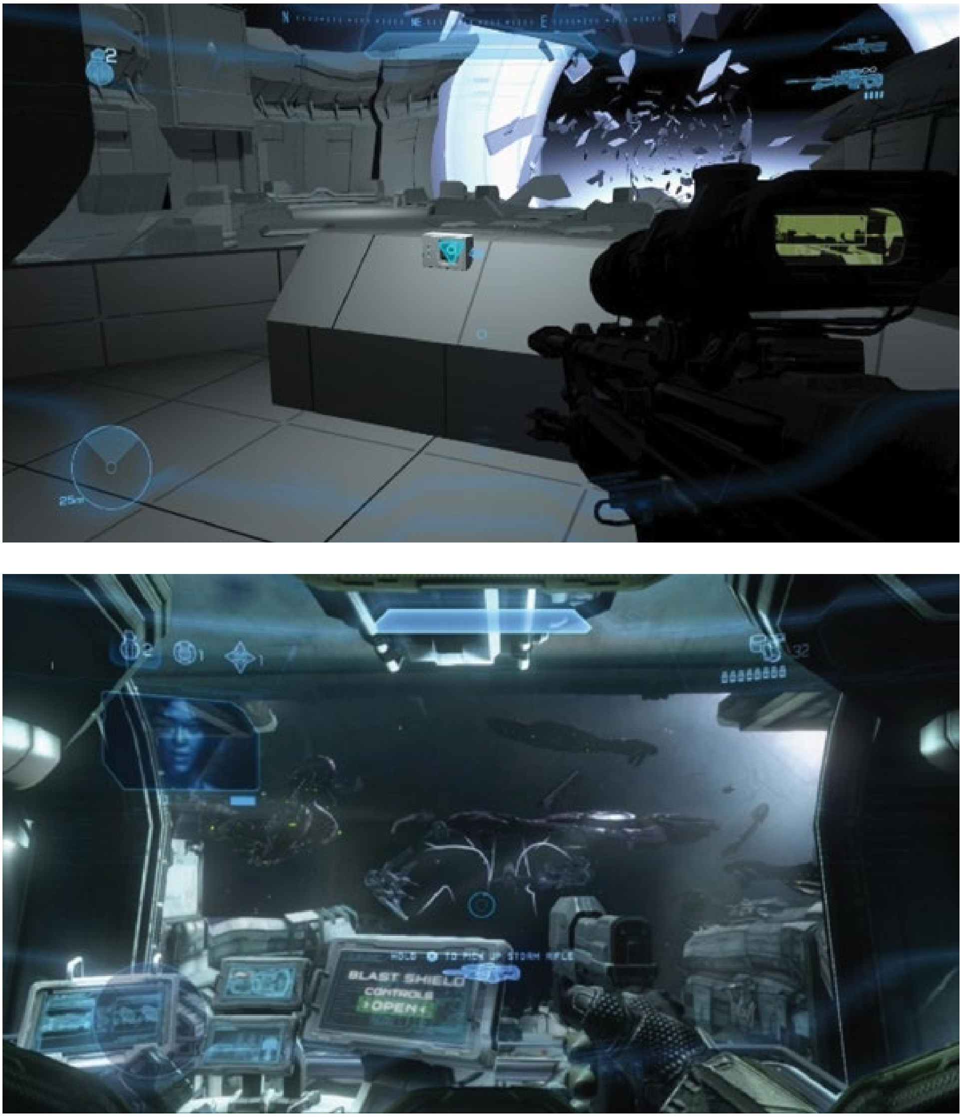


Figure 16.4 At the top is when we first tested gameplay, at the bottom the same scene when we last tested gameplay.

contrast, at the time of the first gameplay test, nothing is typically in the game for narrative. This means that narrative research starts much later and a research schedule has tended to look like Figure 16.5.

The problem with this schedule is that even if you can correctly identify problems in the narrative, it's often too late (or too expensive) to make changes.

The solution

The solution (recommended by Hendersen, 2014a, 2014b) was to stop trying to test the gameplay build and instead to look at the tools the writers were using to ensure that they ended up with a good story. Narrative designers engage as early game designers;

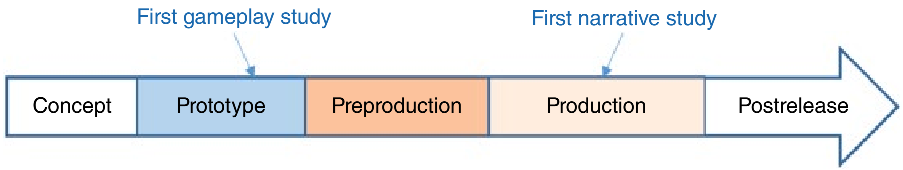


Figure 16.5 Timing of typical first gameplay vs. first narrative study within the game development cycle.

they just prototype their ideas in a different way. Early in the production cycle, narrative designers typically build out a set of narrative beats—the key moments of the game's story. For AAA titles the narrative beats tend to run about three pages, and while they are filled with lots of place holders (character names, "gizmos" that solve locked doors, locations, etc.), they include the logic of the narrative and how it will unfold, from beginning to end. What narrative beats are missing is how the story will be told, also known as the manner of the telling. While the way a narrative is told is very important, one of the keystones of narratology is that a story is separate from its telling.

Originally posited by Benveniste (1971; Chatman, 1978), it is now generally accepted that one can distinguish between a "story" (the event or sequence of events) and the "narrative discourse" (the representation of those events) (see Genette, 1988; or Abbott, 2002 for primers). Consider a fairytale like Cinderella—from Grimm to Disney the way the fairytale has been told has varied (the discourse), but there's nonetheless something consistent across all of the telling (the story). It's the "story" or sequence of events that the narrative beats include, while the "narrative discourse" is what blocks the story from appearing in the game early on. By testing the story, rather than the discourse, we can test early, iterate, and make changes in advance of the expensive discourse being built. Similar to gameplay research early in development, we can use observational think-aloud research methods like usability to understand the player experience.

To test the story the narrative beats need to be converted into stimuli. At times teams have storyboards that are designed to really try to capture the experience—an example from Ryse: Son of Rome (2013) can be seen in Figure 16.6.

The push for these sorts of reified stimuli usually stems from a desire to get at whether or not the story is "good"—however when we are testing the story, the question isn't whether or not the story is likeable. The question is whether the story is understandable. Does the logic of the story flow? Are the characters internally consistent? Are the twists in the story surprising, but does their eventual explanation also make sense? It's far more effective to simply use the beats themselves, compiled into a PowerPoint deck using the following rules:

- Tell the story from beginning to end—do not provide elaborate backstory unless that will be in the players' direct path.
- Use descriptive, rather than interpretive language—so say Character 1 yelled at Character 2 and punched a wall, rather than Character 1 was angry with Character 2. Essentially try to build out a description of what will be on the screen or told via audio and only include that.
- Break before surprises or solutions are revealed—it's very telling to have participants predict in these moments what will happen next.
- Keep it high level—it takes a long time to get through an entire story, talking through each beat.

Figure 16.6 Testing with this stimuli resulted in participants reading all sorts of things into the images, and drawing erroneous conclusions.

We will frequently also include a single picture on a slide, either concept art or a representative still from another game or movie. This gives players something to look at, and helps give a sense of the world, but without distracting them from the facts of the story. Making this deck can be relatively time consuming, but the feedback we've gotten from designers is that the process itself is quite useful:

> As an aside, stripping out interpretive language is actually a good story exercise. Character is revealed through action. So I'm predicting that as I go through this PowerPoint I'll discover that I suck and I need to have some better character-actions in order to have a better story (J. Paquette, personal communication, November 12, 2012).

The feedback we are able to collect from players—by having them read through, explain what's going on, what doesn't make sense, and what they think will happen next—identifies issues with genre knowledge interactions, franchise knowledge (or lack of knowledge), character motivations and inconsistencies, not to mention the basics of plot holes, red herrings, and reversals.

Once a story hangs together for players, a development team and a user researcher can start thinking about how it can be told. Here large-scale playtesting where emotional questions can be asked works well—though it does require that there be narrative in the game to test. While waiting for final cinematics, it's often useful to ask for animatics and temporary voice over. With voice it's always better to use a human over

Figure 16.7 This is at the high end for the quality needed for animatics.

a robot, which is at the very least gender matched to the character and does not change too much over the course of the game. With temp videos, rough flip-book-style pencil sketches work well (see Figure 16.7).

With these things a games user researcher can begin to look at the pace, understanding, and liking of the story, and can begin to build out character sketches. Characters are unusual in that it can be a good thing for them to be scary, or unpleasant, or dumb, and it can be a bad thing if they are heroic, or nice, or liked simply because they are the main character. When building out assessments for characters, it's often useful to use qualitative and nonparametric measures to ask why players like or dislike a character, or how they would choose to describe them (either to a friend, or from this list of adjectives). Qualitative answers will not only tell you how a character is perceived but which elements of the discourse are having the largest effect on perception of the character. For instance, fans of Cortana will frequently quote back lines of hers because here character is expressed through dialog, and fans of Master Chief will describe his heroic behavior, frequently the actions depicted in cut scenes. What success or failure looks like depends heavily on design intent, and it's important to keep the unique and specific concerns of characters in mind.

To know if you've made a successful narrative, there are a few markers to look for in your data. First, players should be able to retell the story from beginning to end. Second, appreciation of the narrative should go up as the narrative progresses. Narrative is a long-burn payoff (unlike gameplay) so it's not atypical for measures of appreciation to be lower in the beginning before gaining momentum, and data curves are often logarithmic as time progresses. Third, once players become highly engaged in a narrative their manner of discussing it often shifts, and instead of simply recounting the facts of a story they will start talking about it in terms of characters ("so-and-so wanted X but then this other person…"), and will also interject comments and opinions in their retellings. Finally, it's important to remember that players appear to have a much higher ceiling when it comes to praising narrative rather than gameplay, and they are much more likely to max out on Likert scales, making it important to benchmark your measures before defining success. Gameplay is great at the excitement, tension, and reward cycle; but narrative can build love.

Case study: Rise of the Tomb Raider

Narrative testing was an early focus for Rise of the Tomb Raider, and we were able to provide player feedback on both understanding and enjoyment of the narrative, leading to several changes. Some of the most substantial changes were to the game's primary antagonists—Ana and Konstantin. Antagonists in heroic stories can be very difficult, particularly when your hero is well known: it's hard to generate threat when most players know that the story will end with Lara winning. Improvements to the antagonists began with structural adjustments focused on "what" the story was and moved onto changes in delivery focused on "how" the story was told.

Testing began early with narrative usability studies, including the slide in Figure 16.8, which shows the original introduction of Ana and Konstantin.

The initial study revealed that participants did not think of the primary antagonists as a major threat, in part because there was no clear explanation of what they were trying to accomplish, other than trying to get the same artifacts that Lara was after. While this set up the necessary gameplay conflict by putting an obstacle in Lara's way, it did not work from a story perspective as Ana and Konstantin just seemed to swoop in and out without having a plan or perspective of their own.

Without a consistent perspective, Konstantin acted more like a miniboss rather than a true villain, and as a result Lara's response to him felt inappropriately strong and inconsistent with participants' understanding of Lara. From the beginning, the story contained a final confrontation between Lara and Konstantin (depicted in-game as a boss battle). However, during usability testing several participants did not think that it made sense for Konstantin to be killed by Lara, something that is problematic for one of the game's primary villains:

> He's a bad guy, he's obviously a villain, you kind of figure they're gonna go toe to toe and he probably won't survive. However, I don't find him to be absolutely horrible, particularly that he's trying to save his sister, which has a redeeming quality to it… However this extreme hatred that she has for him, that she's like "Ahhhh I'll kill you!" and then ends up doing it, seems kind of out of place, a little bit.

- The blonde man takes Lara to the leader of this operation, a woman with white hair and a sickly pallor. Lara learns the names of her captors – Konstantin and Ana.
- Ana interrogates Lara to see what she knows. Lara mentions 'Trinity' and 'Atlas'. Ana suspects that Lara knows nothing and orders her imprisoned.

- Lara is thrown in prison by Konstantin. As he leaves, Lara promises him that she will kill him. He smiles and gives her a small bow.
- Lara explores her cell, looking for any means of escape, when a voice in the darkness in the cell next to her surprises her. She peers into the dim cell and sees a man sitting in the corner – thin, disheveled and bruised.
- He introduces himself as Jacob, and says that he comes from a small village nearby. When Lara tells him she plans to escape, he tells her that he knows the surrounding land and she will need his help.

Figure 16.8 Slide used in narrative usability test of Rise of the Tomb Raider.

Finally, even when Trinity (the organization Konstantin and Ana were a part of) was effective at stopping Lara reaching key artifacts, it didn't necessarily mean that Trinity registered as competent within the narrative. Instead, it seemed illogical for Lara to be consistently blocked by Trinity forces, who would get to these artifacts before Lara, when Lara was the only one who had the knowledge necessary to find said artifacts. Instead of making Trinity seem like a serious threat, this pattern broke the narrative.

This feedback led to three sets of substantial changes:

- First, Ana was introduced earlier as Lara's therapist, setting up a more personal struggle between the two. Later in production, Ana shifted again to being Lara's stepmom, further solidifying their connection and amplifying the sense of betrayal. By combining these roles, Ana became a much more coherent character because more of her history was established and she had a direct connection to Lara, while minimizing the number of characters players had to track and the number of scenes that would need to be built.
- Second, new scenes directly connected Ana, Konstantin, and the paramilitary forces Lara encountered, ensuring that they were understood as a single (substantial and violent) obstacle to Lara's goals. Also, additional scenes were added that made it clearer what Konstantin wanted, and why it was important that he needed to be stopped, however possible.
- Finally, events in the story were changed so that either Lara found the artifact before Trinity, or it was clear that she needed to get to it before Trinity did because they had also discovered the artifact's location. By making Trinity less objectively successful, but more narratively coherent, the threat of the organization increased. The addition of complications for both Lara and Trinity also made events feel more realistic and less contrived.

As production continued, testing shifted from usability to playtests, and we ran a number of playtests specifically focusing on narrative (rather than gameplay). During these tests, participants were encouraged to focus on the main campaign and ignore side content, and test moderators freely offered advice on how to get through potentially tricky gameplay sections. This allowed more participants to see more story content in the testing time allotted, and allowed us to assess opinions of the narrative as well as understanding. For these tests, pencil animatics (see Figure 16.9 for example) and placeholder voice work were frequently used in place of fully rendered cutscenes.

During these tests, it became clear that Konstantin and Trinity were still registering as clichéd cartoon villains, rather than well-rounded characters. While the changes from usability gave players a better sense of who these characters were, and established why they were a threat, Konstantin continued to lack a persuasive delivery.

Again, the production team looked for ways to provide additional context to the characters. They added additional scenes with Konstantin to flesh out his personality (rather than just his actions or function in the story). They also added audio logs scattered through the game world that gave insight into his background and his interactions with Ana. Finally, the team opted to rerecord Konstantin's dialogue with a new actor to more closely match the presentation they were trying to achieve. As testing continued, this led to an increase in the number of participants who viewed Konstantin as a high-quality character.

During the time when we worked on Rise of the Tomb Raider, our team conducted two narrative usability studies and three narrative-focused playtests. In addition,

Figure 16.9 Pencil animatic used in narrative playtest of Rise of the Tomb Raider.

questions about story and character were key components of most of our gameplay-focused playtests. When combined with the testing that the Crystal Dynamics team conducted internally, this research helped to craft a narrative experience that players would find more understandable and enjoyable, and was validated not just by player and critical response, but also by the fact that Rise of the Tomb Raider won the 2016 Writer's Guild award for Best Writing in a Video Game.

Lessons learned

1 **Understanding precedes liking.** Just as with gameplay, if you don't understand the narrative you can't love it. Confusion is not the same as mystery! Only once players can follow the story can they begin to have reliable opinions about it.
2 **A good plot is like a good UI—once it's working well, it becomes nearly invisible.** This invisibility, of course, is why narrative designers will often object that plot isn't the *point* of the narrative—that the characters are far more important. But, until a plot is working well, you can't get to characters—and once it is flowing smoothly, that's all participants will talk about (for good or for ill).
3 **Don't mistake the build for the game.** Early builds are prototypes, and there are lots of ways to prototype different elements of the game. Paper prototypes for UI, white-boxed levels, and PowerPoint renditions of narrative are all ways to test early.

Of course narrative is only one important part of the player experience, and many different methods are required to provide a holistic understanding of the entire player experience with a game. In the next section, we will discuss design intelligence—a way to efficiently understand and act on player behavior.

Providing Design Intelligence

There are many ways to collect feedback from users playing games. Two of the most common include: (a) simply asking them, as you might in a focus group, usability test, or questionnaire; and (b) observing and recording their behavior, as you would in a usability test. Both of these approaches are very common in games user research. They are relatively easy to execute, and provide useful information to inform product development. They are also time consuming, especially coding behavior during usability testing, and thus are most suited to collecting feedback about focused aspects of a game from a small sample of players.

If the goal is to accurately track behaviors and understand what players do in a game, both self-report and observational methods are limited in the accuracy and fidelity of information collected. There are several reasons for this. First, we can't expect an observer to record activity at the rate it is generated when playing a game. Especially in the case of fast-paced action games where players are generating multiple actions per second. Second, games are typically long, sometimes taking upwards of 40 hours to complete. A typical usability test may last a couple of hours. We can't reasonably expect humans to observe and record that much data. Lastly, human recording of behaviors is susceptible to mistakes, even more so in the case of games where the activities are typically challenging, and actions unfold very quickly. For these reasons, when the testing goal is accurate information about player behaviors in games, we typically employ automated telemetry systems and design intelligence.

Telemetry

Telemetry is the automated process of collecting data from a remote application. In the case of games user research and design intelligence, the application is a game, and we are interested in players' behaviors (e.g., button presses), and the game states those behaviors activate (e.g., death, leveling up, reloading). As an example, we may integrate telemetry systems into a racing game that monitors and records every behavior the player makes when driving around a track. Whenever the player touches the gas, the brake, or steers the car, the game records those actions and sends them from the game to a data-collection system. The game will also record and transmit information about the state of the car, such as how much damage it has taken, how fast it is moving, its exact location the track, how many laps the car has completed, and the car's position in the race.

Telemetry is not unique to games and has been around since the early days of behaviorist psychological research methods (Skinner, 1932, 1938). In more modern day applications, such as Formula 1 racing, sensors are built into cars to generate telemetry about the its performance. This performance data is monitored in real time to help the driver control the car as well as inform the pit crew about actions they need to take when the car stops for service.

Almost everything we do using technology leaves a behavioral trace in telemetry systems that organizations use to modify and improve their products and services. Cars record information about how we drive, alerting us when there are issues or it is time for service. In the case of emergencies, car systems may call for help on your behalf. Most of the apps we use on our phones record our logins, activities, and location. For example, most phones automatically tag photos taken with users' location information. Typically, this data is sent back to the app developers to help them improve the service for their customers. Similarly, most websites record what you click, type, and areas of the web page that you hover over. They use this information to build a profile of users so that they can make better product recommendations, learn about their users, and adjust the prices of products. In fact, it is becoming less common to find consumer software or electronics that do not employ telemetry of some sort in the development of products or management of services.

Design intelligence (DI) versus business intelligence (BI)

Throughout a product's lifecycle there are many uses for the data collected through telemetry systems. Two of the most common are business intelligence and design intelligence. Business intelligence, as the name suggests, is the practice of using customer data to make business decisions that, typically, are focused on optimizing revenue. The data may be used to understand what happened in the past (e.g., daily average users), the present (e.g., real-time usage and purchase data), or the future (e.g., predicting how a new product will sell). Design intelligence, on the other hand, is the use of data to optimize the experience of a product or service while it is *in development*. Although there is overlap in the data used for design intelligence and business intelligence, the timing of the analysis, the types of analysis, and the goals of the analysis differ significantly. As the explicit separation of design intelligence from business intelligence is, as far as we know, relatively uncommon, we'll highlight the differences in more detail below.

Timing The timing of these two practices is different. The use of telemetry for business intelligence typically kicks in when a game is made available to customers—BI is primarily focused on product usage after it "goes to market" and the design of the product is mostly complete. Design intelligence begins as soon as a game is playable, often in the very early prototype stage. At times, products may be released in a minimally viable form and further developed using data collected from customers using the product "in the wild." In these case as well, as long as consumer data is being collected primarily in the service of product design (i.e., with the goal of applying it to improve the game), then it is design intelligence.

Data audience The audience and consumers of data differ. The audience for business intelligence is those responsible for making business decisions, such as product pricing, marketing spend, as well those doing portfolio planning for future products and services. Design intelligence, on the other hand, is data provided primarily to product "creatives." In the case of game development, this would include game designers, UI designers, audio designers, and artists, and anyone who is involved in the details of crafting the player experience.

Data size The size and velocity of data differs significantly. Business intelligence data is typically collected from the entire population of customers. A company needs to know exactly how many sales there are or how many users they have. Therefore, depending on the popularity of the product or service, the population can be very, very large and the amount of data can be very, very big. Design intelligence, on the other hand, typically works with smaller samples of data collected from participants during playtests or in product alphas / betas. A typical playtest may have 25–50 participants, depending on the research questions the test is designed to answer. In most cases, design intelligence does not have to deal with data on the same scale or velocity that business intelligence typically uses.

Deliverables The deliverables from design intelligence differ significantly from business intelligence. Business intelligence is primarily involved in postrelease analysis with a focus on high-level usage metrics (e.g., number of people who complete the game). The opportunities for significant product impact are limited, both in terms of the timing of the analysis and the types of analysis conducted. Design intelligence, on the other hand, begins very early in product development and contributes significantly to iteration on game design and the player experience. These two broad goals lead to very different outcomes from design intelligence and business intelligence work.

Business intelligence deliverables are typically an accurate and reliable set of metrics that are referred to as key performance indicators (KPIs). This, in most cases, is a comprehensive dashboard that tracks KPIs against business goals. Daily active users are one such KPI, as are conversion rates (the percentage of customers that purchase something), retention (the percentage of customers that return to the product within a period of time), and so on. In contrast, design intelligence deliverables are typically ad hoc analysis that dive deep into the data in an attempt to answer both the "what" types of questions ("What do players do in situation X?") as well as "why" types of questions ("Why do players do that?") using converging data from telemetry as well as other feedback systems (e.g., questionnaires, interviews).

Typically, KPIs in the absence of converging data streams are of minimal use to those responsible for designing a game. There may be general information about what is happening, but there would be very little information about why that behavior is happening or what the customers think about their experience. Understanding "the what" and "the why" of player behavior are essential to making design decisions that lead to real improvements. The situation would be similar to running a usability study without using a think-aloud protocol, postsession participant review techniques, or surveys. In the absence of those additional data streams, the cause of the behavior observed and knowledge about how it is affecting the experience are up to the researcher to infer with very little relevant data.

Design intelligence at Studios User Research

The first comprehensive use of telemetry for design intelligence at Studios User Research was in 2003 with a game called Voodoo Vince, developed by Beep Industries (Schuh et al., 2008). Voodoo Vince was a platformer with a voodoo doll named Vince. Most of the problems facing the developer were typical for platformer games. We needed to ensure that the approximately 10 hours of gameplay were appropriately balanced and that there was nothing blocking players from progressing through to the end.

On sections of the game we employed standard User Research techniques, such as usability testing. But those are time consuming, and the resources did not exist to observe enough players through the entire 10 hour experience. In the time it would take to do that, the game development cycle would nearly be complete and there would not be time to implement any changes based on the usability tests.

Another option was playtesting. As discussed earlier in the chapter, playtesting involves bringing many players at the same time, and having them self-report using a questionnaire on difficulty and areas where they had issues progressing. The problem with playtesting, however, is that while it is great for collecting information on attitudes, humans are notoriously bad at accurately self-reporting on their own behavior. Further, we needed gameplay information at a granularity that would be impossible for players to report as they played. Having to constantly stop to write down all the little things they are doing in the game is impractical and would have significantly altered their gameplay experience as well.

The solution was to instrument the game with telemetry code that would automatically record what was happening. At the time, we were interested in how long it took players to complete levels, and the location in those levels where their character was dying. With just a few events, we were able to derive accurate data on which puzzles and levels players were finding difficult, and where in those levels the difficulties occurred. Using the data generated from our playtests, we were able to make many design changes to reduce the number of deaths, and to make the puzzles more appropriately challenging.

Following the release of Voodoo Vince we created a toolset that enabled us to more easily instrument other titles, combine multiple data streams together, and get a truly holistic view of the player experience—we called this our TRUE instrumentation solution (Kim, 2008; Schuh et al., 2008). In addition to the tools, we created a group within Studios User Research whose role was to maintain and improve our TRUE toolset as well as provide support for its integration into games in development.

It turned out that while telemetry data was great for helping us understand what players were doing, in many cases we still need to know *why* something is happening, as discussed earlier in the context of KPIs. For example, we can see that players may be taking longer than intended on a level, or that they are dying frequently in an area of the game, but we need to understand why that is happening in order to design an appropriate fix to the problem.

Synching multiple data sources helped us solve that problem. We collect attitudinal survey data at intervals during the session using survey tool running on a PC beside the game at the players' station, we use in-game popup questionnaires that appear when specific events occur, and we record video of players' gameplay that we sync to the telemetry data. When the telemetry data directs us to a potential problem (e.g., many deaths in a mission), we can directly link the death event to the location in the video where it occurred so that we can view them easily.

Design intelligence process

Design intelligence starts very early in the product life cycle, and continues until game iteration stops. Below are the typical design intelligence activities during the early, mid, and late stages of game development.

Early Early in product development the design intelligence experts, using product documentation and early prototypes (sometimes referred to as a "vertical slice"), develop use cases for the research they plan to conduct on the title. For example, a high-level use case may be written as "What areas of the level are difficult for players?" This question is then refined to more specific questions such as "Where on the map do players below level 3 experience difficulty battling alien grunts?" With this level of detail, we know we will need to record player location, character level, and the type of enemy fought. At this early stage, there will likely be several functional groups developing their own use cases, and telemetry requirements.

Mid As the game develops, and it reaches a state where it can be played by customers, the design intelligence work kicks into full gear. With the game telemetry systems in place and our TRUE system integrated, it is time to collect usage data from real customers. At this stage we bring groups of players into our playtest labs, and they will play the game for a specified amount of time, or however much content we are testing. The primary goals at this stage are: (a) to determine whether player behaviors are consistent with how the designers intend the game to be played; and (b) to start collecting feedback on game difficulty (both in terms of learning systems as well as the challenge of the game). We also employ usability testing and other evaluation methods concurrently with telemetry to provide multiple sources of consumer play data for game development.

Late In the later stages of game development, when most of the content is in the game and playable, design intelligence will focus on: (a) validating any design changes that have been made based on prior issues found; (b) checking that no new issues were introduced; and (c) providing a holistic evaluation of the game now that most of the content is testable.

Another important function of design intelligence at this stage is to prepare for the business intelligence work that is beginning to ramp up as the game nears release. If design intelligence has done its job, the team will have built up a considerable amount of information about player behavior that will be relevant and useful to the business intelligence team as they track metrics important to the product in the wild. For example, design intelligence may have identified factors likely to be important to retaining customers for the product. This can then be tracked by business intelligence when the game goes live to aid in retention prediction metrics. Design intelligence will also have spent months adding events, and refining the telemetry system and data pipeline, which will be employed by the business intelligence team postship.

Lessons learned

We've been doing design intelligence since 2003 at Microsoft Studios. We've had a lot of success improving products, but we've also had to struggle and refine our tools and processes to make them work in a fast-paced game-production environment. We've learned a few things along the way that we share with you now.

Messy data Despite what is taught in university statistics classes, human behavior is complicated and messy, and outliers are frequent rather than the exception (which might be an oxymoron!). The vast majority of data is generated from behavior that is skewed, and it is common for it to contain more than one mode. All of which makes it difficult to summarize what a typical case is using averages, and it is important to apply the appropriate statistical tests when making comparisons within or between tests. We've learned that it is always necessary to visualize distributions of data before trying to summarize them.

Telemetry data collected from gameplay often contains outliers—cases where a few players generate data that differ significantly from the rest. While these are typically treated as exceptional and removed in the practice of business intelligence, they are extremely important in design intelligence as they likely represent activities or player types that should be accommodated for in your game design. It is possible that the player who died an excessive number of times getting through a level has just found a bug or game pattern that needs to be fixed.

Counting things The second thing we've learned and would like to emphasize is that simply counting things will get you 90% of the way. Don't be scared by talk of data science and machine learning, and all the sophisticated tools you can possibly use to analyze data. Over the years we have used sophisticated statistical methods, but they are a very small minority of the analysis. The majority of product improvements based on design intelligence have come from simply counting the things we are interested in.

Emergent experiences Emergent experiences are composed of discrete activities, but are different than the sum of their parts. The identification and analysis of these experiences, which is difficult, are important for doing good design intelligence. Consider this analogy. A restaurant owner wants to improve the customer experience in their restaurant. To that end, through use of magical telemetry they are able track every time customers use a knife, spoon, glass, napkin, and all the various tableware. They have a record of every discrete action each customer took during their visit to the restaurant. But knowing the number of fork or knife uses or how many sips of wine a customer had is not sufficient for understanding the experience of eating. It will give you general usage data, like that typically tracked by business intelligence, but it won't give you much insight into the dining *experience*. The experience is emergent from the discrete activities constituting it, and is more than the sum of its parts. The data only makes sense in the context of other events as well as knowledge we gain by asking the user about their experience. Identifying emergent experiences is difficult and requires an analysis of events in sequence and within the context of converging data sources.

Targets The last thing we'd like to share is the importance of using targets. Data, on its own, is ambiguous. Are 10 deaths for a level in a shooter game a bad thing, an ok thing, or a really good thing? It all depends on what the design intention is. In order to get to where you want to go, you need to know where you are headed. Without having targets for the behaviors and experiences you track, you have no way of knowing whether you are making any progress towards your goal. Just as we emphasized earlier in the chapter, in order to be able to test the player experience against the design intent, games user researchers must have a solid understanding of the design intent. Effective design intelligence requires a close relationship with design teams, and a very good understanding of the experience they are trying to create.

Conducting User Research with Kids

Conducting User Research on children (<12 years of age) raises many challenges that we don't encounter when conducting research on adult participants. In this section, we outline several of the obstacles that we've faced over the years, as well as our solutions. To be sure, these aren't the only viable solutions to these problems—these are simply solutions that have worked for us.

Study logistics

Recruiting Most User Research labs have an established mechanism for recruiting their target audience, which is typically adult consumers meeting some specific criteria. But there are several challenges with recruiting the junior audience.

There are many methods to reach out to potential adult participants—social media, targeted advertisements, email, word of mouth, etc. But reaching the younger audience is not as simple. We can't contact a 5 year old via email. And we don't anticipate many 7 year olds visiting a Facebook page or Twitter feed. To recruit a substantial number of children for our research, we had to rely on unconventional methods: partnering with local preschools and daycares by dropping off flyers and brochures in parent folders; setting up tables at those centers to speak directly to parents and address any concerns they might have; sending staff to family-targeted events to set up recruiting booths. This required a lot more effort than recruiting adult participants, but we were able to fulfill our research needs successfully in the end.

Timing Once the younger audience has been recruited, the next issue to arise is finding a time to conduct the research. Kids-related research studies take a lot longer to run simply because it is difficult to schedule the studies. With kids in school all day, and often booked with after-school activities, available hours are limited. With that, we found that we were forced to run sessions on weekends as well as during the late afternoon / predinner hours. We also plan to run much of our research during summer break, although product cycles don't necessarily align with that schedule. One other solution was to take note of all of the school holidays and teacher in-service days for schools in the area, and attempt to use those days as data-collection opportunities.

Staff With adult participants, staffing requirements are minimal. Most of the time, a usability study can be conducted by a single researcher. With children, the staff requirements double or even triple. For much of our research, it is unreasonable to leave a child in a room by herself while the researcher conducts the session from behind a one-way mirror. We found that it worked best to have one researcher behind the mirror taking notes and staffing the cameras and tech, while a second researcher worked directly with the child in the participant room, giving verbal instructions and demonstrations where appropriate. In addition, we sometimes required a third researcher to sit with the parent and explain the session, as well as keep them from interfering (see below).

Methods

Building rapport Working with children differs from working with adults in many ways. As discussed above, we are not able to simply leave them in a room by themselves with a task list and expect them to think aloud while playing a game. Generally, we need to have at least one researcher in the room with the participant to help them navigate through the study. It is of critical importance that the researcher is well-versed in working with children. This not only means being able to develop a good rapport with the child, but also being able to convey the tasks appropriately so that the participant knows what is expected of her.

When working with any participants in a 1 : 1 situation, it is important that the participant feels at ease with the researcher (Nielsen, 1993). With kids, rapport building is critical. This allows them to speak more freely and behave more naturally, providing much better data than if they were uncomfortable. To build rapport with our participants, the researcher took time to sit with the child to complete some completely unrelated tasks. For example, they might complete a puzzle together, draw pictures, or play a simple game. This helped get the children familiar with the researcher, and made them feel at ease during the session.

Parental interference When children are in the lab, it is important that their parents always have a visual line of sight to them. This not only gives the parent extra peace of mind, but also protects the researcher should anything out of the ordinary happen. However, in early testing, we found that parents were often too eager to jump in and help their children on tasks at which they were failing. Or parents would verbalize for their children with the intent of being helpful, but with the outcome of depriving us from hearing directly from the child. In some instances, parents would bias their children's subjective responses by trying to influence their answers in an effort to please the researcher. With all of that, we devised a set of solutions that helped maintain the line of sight, and provide the close proximity between parent and child, while minimizing the likelihood of parental interference (see Figure 16.10). Using the traditional usability lab's one-way mirror, we positioned the parent with an occluding screen directly between her and the child. But we maintained line of sight by asking the parent to look at their child in the reflection of the mirror. With the way that we positioned the stimuli and participant, this let them see the backs of their children while they participated in the study, and minimized interference or participant distractions.

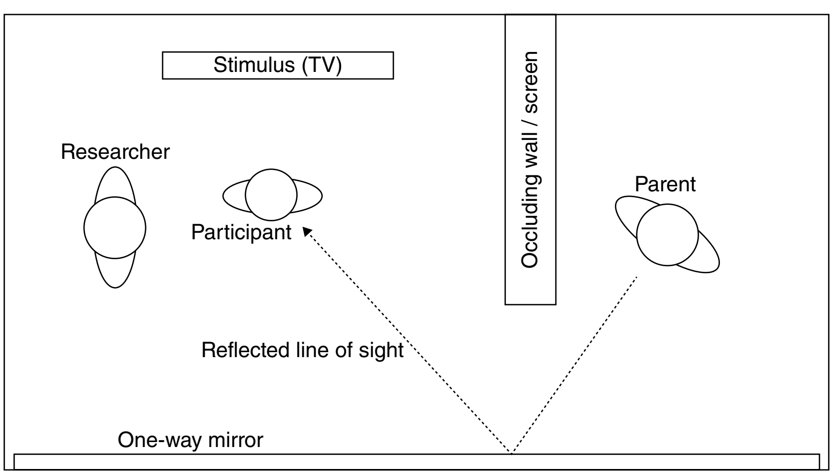


Figure 16.10 Lab setup to maintain parental line of sight but minimize interference.

In addition to these measures, we sometimes ask the parents to fill out surveys or other forms while their children are participating. This not only further reduces interference, as they are distracted by the paperwork, but also allows us to collect some additional data.

Gathering subjective data Getting children to tell you how they really feel can be a challenge. It is common for children to tell the researcher that they love the experience simply because they are being given a researcher's undivided attention for a period of time or they're currently engaged in a completely novel experience. For adult-targeted gaming experiences, researchers generally quantify the subjective experience based solely on participant self-report. With younger participants, we have to use a variety of inputs to accurately convey the subjective experience to product teams. For any given study, we might rely on rating scales, card sorts, observer ratings, comparisons to baseline experiences, or spontaneous verbal responses, depending on our capacity and goals.

We've used simplified rating scales with happy, neutral, and sad faces (see Figure 16.11) for kids to tell us how much they like an experience or visual asset. We found that kids were generally good about using all three points on the scale, and, from what we could tell, giving a response that matched their feelings.

We've also used card sort tasks, using boxes labeled with a happy and sad face. For these tasks, we gave kids cards that represent a character or visual asset, and asked them to put it in the box that represents how they felt about it. Often, we add some warm-up tasks, asking kids to sort items that we've already established an affinity for or against. For example, sort cards that represent their favorite and least favorite foods. These warm-up tasks help prime the participant to let them know that it is fine to put something in the "sad / dislike" box.

For some studies we used observer rating scales to try to assess how engaged and interactive participants were with a given experience. In these studies, we were testing an interactive television show, which combined linear content with moments of interaction. We had trained observers monitor and record how engaged the participants were during key moments in the show.

Along those lines, when trying to measure engagement and fun with a Kinect experience, we used a baseline comparison to determine if the in-development product was on the right track. For example, we knew that our young participants enjoyed seeing their own Kinect figure on the screen. The Kinect development system contained a video feed of the skeletal data it was collecting, which was essentially a stick figure representation of the participant on a black screen. There were times when we used this experience as a baseline to which we compared the target experience.

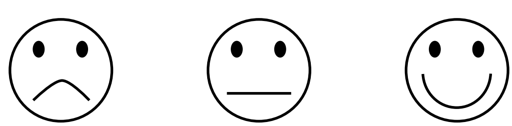


Figure 16.11 Adapted rating scales for kids.

Case study

In User Research, we can sometimes be surprised by data. It could be an unexpected score, a larger than anticipated problem, or an unusual observation in the lab. But with User Research on children, the atypical becomes the typical. It is important for product teams to be open to the likelihood that children will not respond to their games in the way that they expect.

With one particular title, we encountered numerous instances where the design intent and the observed behaviors were completely mismatched. As an example, we tested a particular Xbox Kinect title that used augmented reality to portray fantastical scenes with the players' video images directly placed in the scenes. Children had the propensity to move closer and closer to the TV. This occurred because they were excited to play the game and loved seeing themselves on TV, and if playing with others, often jockeyed for position to be in front and more prominently on camera. The problem occurred when players were too close to the Kinect camera. It had trouble picking up their images and wouldn't function as designed.

In an attempt to discourage kids from charging towards the TV and camera, the designers built a character that would show up on the screen when somebody was too close, and disappear when they were back in the correct position. In a playful manner, the character would gesture for the player to step back to an appropriate distance. When we tested this new mechanism, we discovered something interesting. Instead of the character acting as a discourager, children perceived the character's appearance and disappearance as being the goal of the experience. They moved forward to make him appear, and back to make him disappear. This became the game itself—peekaboo with a video game character. Clearly the attempted solution was a failure. We then analyzed the problem further to figure out a different solution. If children were charging the screen to better see themselves on TV, then perhaps the solution was to take away the reward for approaching—seeing themselves on TV. Now when children got too close to the TV and camera, their image would blur significantly, to the point that it was unrecognizable. The only way to get back into focus was to step back to an appropriate distance. Upon testing this solution, we found that it worked exactly as intended. Children would approach the TV and camera as expected, see that their image became unrecognizable, and step back to become visible again.

Lessons learned

Testing younger audiences requires more time and resources than normal. When testing this audience, recognize that recruiting and scheduling are far more time consuming and difficult with kids than adult participants. Keep in mind, too, that conducting the research requires additional staffing and space.

Careful planning of the testing sessions is critical for success. Spend extra time identifying potential hurdles that may arise.

When testing kids, the unexpected happens more frequently than testing with adults. No matter how prepared you are, there will often be events that were never anticipated. An overly engaged parent, a jealous sibling, a shy child, or an extra three bathroom breaks—this is a small sample of the types of hitches that can occur when testing kids. Don't expect a study to be run without incident. Instead, be prepared to be flexible in testing methodologies, and in data gathering.

Understanding User-Generated Content

User-generated content, or UGC, is becoming increasingly prevalent in games. This is content that is created by the community, often with the intent of sharing with others in the community. For example, in games like Minecraft, players create massively detailed worlds and structures, and share those out to fellow players. In Forza Motorsport, some players spend countless hours designing paint jobs for their virtual vehicles that other racers can see and use for themselves if desired. In Super Mario Maker, players unlock editing tools and create levels that others can play for themselves. UGC gives players the opportunity to be creative. It lets them share their skills with others. And it lets people take and iterate upon someone else's work to come up with something brand new. When done right, UGC systems have the ability to promote sustained engagement with a game, leading to a dedicated audience that plays the game longer than it would have if simply relying on content built by the developer.

The builder's journey

Through our research on the UGC community, we have learned a lot about what motivates them, how they progressed as content creators, and pitfalls that they've encountered along the way. We identified something that we've dubbed "the builder's journey." This framework is an idealized version of the stages through which a user would progress in order to become an advanced content creator.

As with many journeys, it starts with the builder learning the tools of the trade. They need to become familiar with the toolset, and know where and how to get help if needed. Toolsets need to be flexible enough, and easy enough to grasp, so that new users can get into the system and start creating with minimal friction. When friction is encountered, the system needs to be supportive and let users get help in a simple way. Occasionally, users will share their creations with a trusted circle—friends, family, and maybe the community with the goal of getting useful feedback on a problem. If a toolset is too cumbersome or clunky, or if users aren't able to get the help that they need, they will often abandon the experience for good.

Once users become proficient with the toolset, they move into the early creation stage. Within this stage, we identified a "contribution cycle." This is the process by which users start creating meaningful content. The contribution cycle is composed of three parts—imitation, creation, and iteration. Often, lower skilled content creators begin with imitation, copying content that they find interesting with the intention of learning and becoming better at the toolset. However, it's not just lower skilled creators who start with imitation. Higher skilled users often imitate in order to find inspiration for their own work. Within this contribution cycle, users will not only imitate, but they will iterate on their creations until they are satisfied with what they've built. In addition to the contribution cycle, at this stage it is important that users are able to utilize content libraries effectively. These are libraries containing UGC from other content creators that users want to browse in order to be inspired and to learn. During this early creation stage, users are often reticent to share their content out with the rest of the community. They worry about negative feedback or their work not being good enough. With that, users want to have control over who they share content with, such as a small group of friends or trusted community members. If users' sharing fears are never resolved or if their content browsing experiences are suboptimal, they may end their journey here.

Finally, if users have become proficient with the tools, and are building and sharing their own creations, they move to the advanced creation stage. Here, users are sharing willingly with the community, getting feedback and recognition as high-quality content creators. Feedback can be explicit, such as upvotes or positive comments, or implicit, such as knowing how many users are downloading their creations. Advanced creators are also valuable community members, providing support and feedback to others in the community. Creators at this stage build because they love it, but if developers use their creations for profit purposes, they expect some type of monetary reward. Advanced creators also have an expectation that the community will be positive and vibrant, a place to share, learn, and get useful feedback. If users aren't rewarded for their work, if the community is stagnant, or if there is a lack of quality content being shared by the rest of the community, even advanced creators will leave the system.

Our research with this population helped us generate a number of insights that we delivered to our teams. It was already realized that a vibrant community of content developers and consumers was important for extending the life of a product, but the path to achieving this goal wasn't well understood. Identifying the stumbling blocks and dropout points through these investigations enabled us to equip teams with strategies for keeping players engaged and motivated to become successful content creators.

Lessons learned

With the more strategic projects that we tackle, it has become more obvious that to really understand the issues and the nuances we need a multiple-pronged approach to our research. Specifically, combining qualitative research techniques such as interviews and field work with quantitative methods like large-scale surveys has helped us develop deep and meaningful insights for our teams.

In trying to understand the UGC world better, there were a lot of delicate topics that arose out of the qualitative interviews, such as admitting a lack of skill or discussions around monetary rewards. Understanding how to navigate these issues to get the most relevant data points is an important skill for a researcher. As we move towards striking a proper balance between quantitative and qualitative methods, it has become increasingly important to have researchers who are practiced in a variety of techniques.

User Research in Motion Gaming

From a design and User Research perspective, motion game development presents just as fascinating a set of user experience challenges as that of traditional games. Three central tenets for good game user experience perhaps apply most strongly for *motion games*—know the person (capabilities, limitations, expectations, and preferences), know the place (physical and social context), and know the product (motion game system and game demands). In each of these areas, motion games tend to be nontraditional.

- **Person:** For a variety of reasons (including input fidelity and responsiveness, preference for seated versus active play, and available game portfolios), the audience for current motion games tends to be younger and / or has little overlap with the audiences for traditional game genres like shooter, action-adventure, RPG, MOBA, etc. Age of player is related to ability to perform motion inputs consistently and predictably, as well as related to the prior knowledge the player brings to the game.
- **Place:** Context refers to both the physical and the social context. A game of Kinect Sports may include multiple players (with potential for unintended physical contact), multiple nonplayers / observers, and well-intentioned helpers (and those less well intentioned!) who may interfere with the detection system. Physical context also must be considered, as these games require varying degrees of space, and the Kinect in particular requires a cleared playspace.
- **Product:** Finally, the technology underlying the system and game is of course nontraditional as well. A motion gaming system that combines gesture, voice, and physical controller may have more flexibility (and greater complexity of user experience) than a purely gestural system.

For the purposes of this section, we will consider motion game development on the following systems—Kinect for Xbox, Nintendo Wii, and Playstation Move. (We expect that, with the recent proliferation of augmented reality and virtual reality game system development, motion gaming will continue to grow, but as many of these systems are still in development, they are beyond the scope of this chapter.) These systems differ in their underlying technology but they share a number of user experience considerations and challenges.

Input system

The most notable characteristic of motion games is of course the input system. We use the terms "action" to refer to the player's input, "activity" to refer to designed game experience (e.g., throwing a ball, swinging a sword), and "input model" to refer to the specific recognition algorithm that assigns correct / incorrect to a player's action. Because action is not constrained the pressing of a button, the player has practically infinite degrees of freedom to make an action. This leads to the problem of input.

For the purposes of understanding and explaining player interaction, we break input into two categories—player intent and player action. These are equally important to consider during development and to attend to during user testing.

Player intent Players' behavior is driven by their prior experience and by the cues (and our interpretation of those cues) in the game. Consider a sword-fighting game.

- **Experience.** The player brings prior experience and knowledge of sword fighting to the game. One player may choose to thrust vigorously toward the sensor, while another interprets the game to require sweeping motions across the body. Whether the game design intends one or both of these to be correct, it is likely the game will need to accommodate both (and probably more) sword fighting models.

- **Cues.** However, as soon as the player sees how the game represents that input, a feedback loop has begun, and the player will use these cues to quickly update their inputs. If the onscreen sword makes sweeping movements even in response to the "thrusts" of the first player, that player will likely begin to adjust her arm movements in response to the onscreen cues. However, the player will likely exhibit a mix of the desired (sweeping arm movements) and original (thrust) inputs.

This example captures the challenge of designing good input models for motion games. While the instinct may be to develop tight, precise models, the degree of variability within and between individuals is so great that this will almost certainly result in failure. We recommend user research to understand the range of player actions your game activity may elicit, and then designing "loose" input models that accommodate a range of approaches to that activity.

Player action Further complicating matters is the fact that players often intend to perform a specific action but actually perform a variant of that action. For example, in Dance Central, it was common for players to think they were performing a move correctly, only to be told by onlookers that their arm wasn't high enough or they weren't stepping far enough. (This example also illustrates the value of setting up user tests that replicate the expected social context of the game.)

There are very good reasons for why we are "bad" at actions in motion games. When we play games, we are carefully attending to what is happening on the screen, not attending, for example, to where our feet are or how high our arm is raised. Player action is highly variable in part because players simply forget to pay attention to where they are in the room and how their body is moving. Often this manifests in minor variability of inputs that the player isn't aware of but can cause confusion if the variability causes decrements in game performance. In our testing, we've seen more egregious examples of this, including:

- **Player drift**—Players moving around the room while playing, at times moving outside of the range of the sensor (a particular problem for the Kinect).
- **Camera creep**—A specific version of drift, where players (especially very young players) moving closer and closer to the screen, until the detection system completely fails.
- **Collisions—**Players will accidentally bump into furniture, tall players may hit overhead lights or fans, and in multiplayer games, they may hit each other. Instances of these have been quite popular on YouTube, but of course these scenarios can be dangerous.
- **Fatigue**—Fatigue can impact the fidelity of player action as well, leading to slower actions and/or more variable actions.

Learning and feedback systems have been developed for most motion gaming systems to help players understand the importance of creating a safe playspace, as well as communicating to players what the playspace is, where it is, and where players are within the playspace.

Instructions, cues, and feedback

We've painted a fairly bleak picture in the prior section. Is player action so variable that developers should give up on hopes of controlled and purposeful interaction in motion games? Not at all, and there are three classes of solutions that can help to guide and constrain player action—instructions, cues, and feedback.

First, though, we must dispel the myth that gestural input doesn't need instruction. An early marketing slogan for Kinect stated, "The only experience you need is life experience." But as we have discussed, there is great variability in the form that different players use for even simple gestures like pushing a button or tossing a ball, and most current motion gaming systems simplify processing and development by strictly defining correct and incorrect action. For example, for an action to be classified as a "throw," the hand may have to reach a certain height above the head and move from a particular point behind the head to a particular point in front of and below the head, moving at a certain speed. If the player chooses to throw in a side-armed or underhanded fashion, a throw is not registered. Multiple valid throws may be defined, but as the gesture space becomes more and more "cluttered" with valid inputs, false positives will begin to creep up, leading to faulty feedback, and players will struggle to learn and begin to distrust the system. Instructions and cues can serve to guide and constrain players' actions, while feedback can provide useful information to help players continue to improve their performance.

Instructions and cues We define instructions as explicit guidance to perform actions in a particular way (often included as part of a tutorial or other learning system). Cues, on the other hand, are auditory or visual elements in the game that provide more tacit guidance. Cues are common in games—the iconic beeps leading to the start of a race in Mario Kart tell us when to press the accelerate button. With gestural input, visual cues can be used to great effect to shape player behavior and encourage players to move in just the particular way that the game requires for a valid action. For example, in Kinect Sports bowling, the player gets a new ball in their avatars hand by reaching toward the line of balls positioned to their left or right. The placement of the balls encourages the player to reach out from their body to a given side, an action that is recognized by the system as the command to generate a ball in the avatar's hand.

For more complex actions or when multiple actions must be used by the player, explicit, visual instructions are valuable. Important characteristics for good visual instructions include:

- Player models in the instructions should be humanoid, even if the character being controlled is not.
- Instructions should include all body parts that are relevant to the input. If the player's stance is relevant, depict the player model's full body. Similarly, relevant limbs should be highlighted to emphasize their importance.
- Player models should be animated to show the full range of motion, including the start and end positions of all moving body parts. Movement should be depicted at the intended speed, as players may emulate slow motion instructions. If fully animated instructions are not feasible, at least show the start and end positions, with arrows to indicate the direction of the motion.

- Player models should be depicted from an angle that clearly shows all relevant parts of the body. For example, a player model directly facing the player may not communicate depth effectively.
- If the required input conflicts with prior knowledge or expectations (either from the player's life experience or earlier experience in the game), be sure to provide instruction. For example, Figure 16.12 shows a train door in a Kinect game. Many players were unfamiliar with the sliding nature of train doors (trains being less common in the United States), and their prior knowledge led them to push or pull. (Note that the small white and black cue was either not noticed or not understood.)
- Players are more likely to mirror the visual model than try to match the model (although younger players may not understand the concept of mirroring versus matching, and may alternate between the two).
- Instructions should be presented at a time when players can attend to them. "Just-in-time" instructions will often be missed or ignored, unless the game pauses to allow players to inspect the player model in the instructions.
- It may be useful to re-present instructions if the game detects that the player is struggling to correctly perform a gesture, as a form of feedback.

Feedback Despite well-designed instructions and cues, players will still sometimes perform the wrong action. They may perform a valid action at the wrong place or time or they may perform an action incorrectly (whether close to the correct action or completely wrong). Well-designed feedback is crucial. Feedback should be consistent and useful.

Some games do not give feedback on incorrect actions, assuming that the player will understand that the lack of change in the game state signals an incorrect action. However, as motion game systems are not as reliable at correctly detecting inputs as physical controller games, players learn that a nonresponsive game could mean either (a) an incorrect action or (b) a failure to detect a correct action by the system (or a "miss"). Thus, feedback should be given for correct actions as well as incorrect (i.e., the game detects an action that does not fit the correct input).

Useful feedback is feedback that gives the player some understanding of what went wrong. Useful feedback does one or more of the following:

- Communicate what body part was out of position
- Provide tips on common ways the action might be performed more accurately (e.g., "remember to raise your hand high above your head!").
- Clearly show what action the player did, allowing the players to draw comparisons between what they saw themselves do onscreen and what earlier instructions depicted.

One way to show the player what action was performed, and one of the most useful feedback systems in many Kinect games, is an always onscreen player avatar. The onscreen avatar provided a constant feedback loop to the player, telling the player what the system sees, as well as telling when the system was not "seeing" the player well. In the Dance Central games, the onscreen character was not a 1 : 1 mapped avatar, but the character representation still effectively communicated which limbs were out of place or out of sync via a highlighting system.

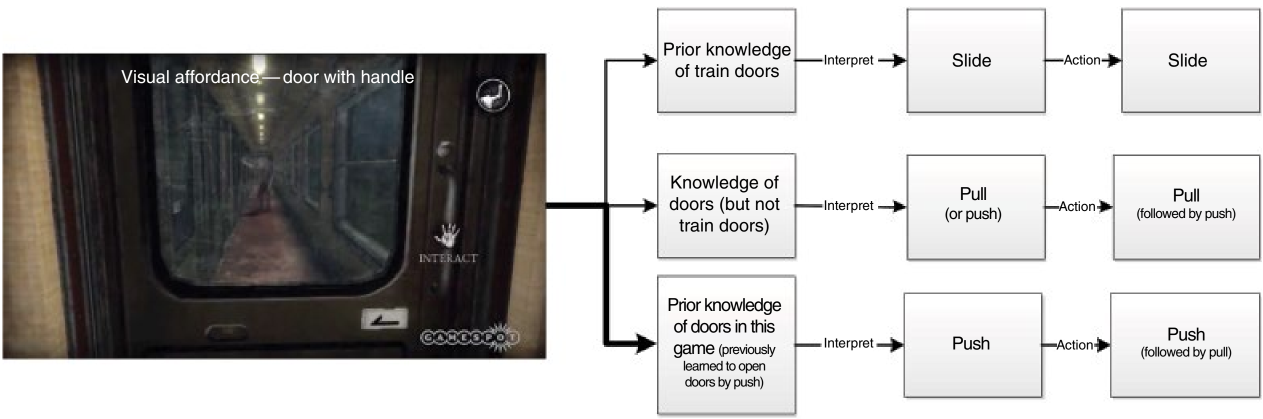


Figure 16.12 Train door in a Kinect game, D4: Dark Dreams Don't Die (2014). Lack of prior knowledge with train doors in some players led them to push / pull instead of slide.

It is important to remember that, unless specifically against your game design, the act of remembering the right gesture, and the right way to perform that gesture is *not* the fun in the game. There is no harm in providing fully corrective feedback to the player when an incorrect gesture is detected, and in fact, there is considerable harm, in the form of player frustration and distrust of the system, in not providing quick, useful feedback.

Lessons learned

- Our overarching recommendation is to assume extra iteration time in developing the input model for each of the activities in the game. Motion controls are harder to get right than physical input devices.
- Variability is a given, both player-to-player variability, and variability within each player.
- While instructions, cues, and feedback can be useful in constraining and guiding player action, these are not a panacea. Motion game designers must build to the strengths of the systems and recognize the limitations. Don't make games for audiences that demand precision and low latency controls—these aren't currently available in motion gaming systems.
- Players are not good judges of their own body position when focused on the things happening onscreen. They often do not attend closely to their limb positions, nor do they correctly perceive timing of actions.
- Avoid strict requirements for correct gestural input. The fun in these games is typically not in trying to remember and perfectly match an action to an ill-defined input requirement. Design loose constraints to allow a range of players be successful.

The Lure of Biometrics

The lure of biometrics in games user research is undeniable. Just as in cognitive psychology, biometric methods offer the promise of insight into the mental state of our users without needing to directly ask them. However, just as in cognitive psychology, the data provided by biometrics in games user research is difficult to collect, messy, and complex. Very frequently, the time and resources required to acquire and analyze biometric data do not produce superior insights to those generated with more traditional (but less flashy) methods. Therefore, it is critical to understand how best to implement these methods. When asking the question, "Do you really need this type of data to make your product better?" the answer is usually "Probably not."

Biometrics refers to the various data streams acquired by taking measurements of certain physiological processes or states that are thought to be indicative of a person's mental processes. Examples include eye tracking and pupilometry, electrodermal activity (EDA), heart rate variability, and electromyography (EMG) of facial muscles.[1]

[1] Techniques such as electroencephalography (EEG) and functional magnetic brain imaging (FMRI) also fall into this category, and have the added appeal of more closely measuring brain states. However, they are even more costly and complicated than the measures listed here. There are some commercial EEG set-ups that are marketed toward User Research scenarios, but even these systems will provide a complex multidimensional data set that can be difficult and time consuming to convert to actionable insights. If anyone ever tries to sell you on using FMRI for user research, firmly grip your wallet and move very quickly in the opposite direction.

Eye tracking is a fairly straightforward measure of gaze direction. Measures such as pupilometry, EDA (commonly referred to as Galvanic Skin Response or GSR), and heart rate variability are typically associated with arousal, engagement, and focus. EMG is thought to give a more direct measurement of actual mood (happiness, frustration, etc.).

It is important to note that all of these methods are *inferential*: certain assumptions need to be made in order to connect the observed phenomenological changes to the mental states of interest. For example, based on established psychological research, we *assume* that an increase in EDA (indicative of increased perspiration) is a marker of increased arousal. However, we are not measuring arousal *directly*, just its byproducts. This is true even for something as relatively straightforward as eye tracking: sensors detect where in a display the pupil is directed, but this does not guarantee that a person is actively *looking* at this part of the display. People can point their eyes at an object and still not see it, as Dan Simons (theinvisiblegorilla.com) and many others have demonstrated. Indeed, even techniques that are used to measure "brain activity" often measure the side effects of mental processes, rather than the processes themselves. Fairclough (2009) provides an extensive discussion of the challenges in matching specific measures (or groups of measures) to specific mental states or psychological responses; although the primary focus of that paper is physiological computing, many of the same issues apply to biometrics in games user research.

If biometrics are multiple layers removed from what we really care about (a person's mental state), then why bother with them at all? We bother because, under the right circumstances, they get us close enough. Not only that, but they get us close without asking someone directly (and relying on interpretation of both question and answer). As Sasse (2008) discusses, surveys and think-aloud paradigms introduce a certain level of removal from the core gameplay experience and can reduce immersion. Observing physiological changes in the players brings us closer to their root experience. Additionally, they bring us closer in something resembling real time. Most of these methods measure processes that vary on the order of seconds or fractions of a second. This granularity is the difference between asking whether someone liked a particular level in your game versus identifying the specific moments in the game that they liked (or didn't like).

Granularity is then the key advantage of biometrics. Because these biological processes have relatively low latency (i.e., the time it takes for a person's mental state to be reflected in their physiological state), it is (theoretically) possible to build a moment-by-moment picture of how a player is feeling as they are playing a game. The researcher would be able to see frustration shift to excitement and joy as a player figures out how to solve a puzzle, or identify the point at which a battle shifts from being engaging to being tedious, as Mirza-Babaei & McAllister (2011) have described.

Figure 16.13 shows another selling point of biometrics: it *looks* cool. Sometimes, the presence of a fancy heat map or EKG-like display can make a team feel like they are getting super fancy scientific analysis in a way that survey questions just don't. Because these visualizations are so compelling, it is also that much more important that care and consideration are taken with biometric data analysis. It is much easier to lead teams to erroneous conclusions with such fancy images (see McCabe & Castel, 2008, for an examination of this issue using brain-scan images).

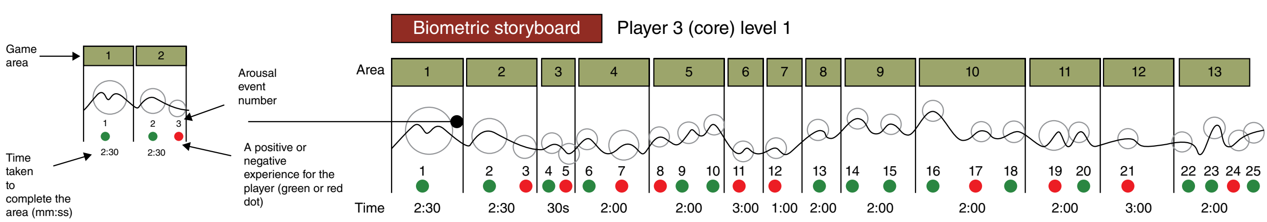


Figure 16.13 An example of biometric storyboard. Taken from Mirza-Babaei & McAllister (2011); used with permission.

The challenge of acquiring data

Data collection for biometrics is not trivial, though. In terms of cost and time commitment, it is nearly the polar opposite of discount usability. All of these methods require special equipment and software to collect the measurements and record the data. The analysis of the data can be time consuming as well and requires a greater level of statistical sophistication than survey and usability methods. Pernice and Nielsen (2009) describe the various considerations in properly collecting eye-tracking data. Figner and Murphy (2010) describe in some detail the setup and analysis of EDA data, as another example; similar guides can be found for other methods. These guides show how biometric methods are technically and analytically complex, and not lightly undertaken.

It should be noted that there are several companies right now that provide off-the-shelf solutions for biometric data collection. These solutions are generally not cheap, but they do provide the software and hardware necessary for data collection, and often provide a straightforward analysis of the results (condition A is good, condition B is bad). Such systems can also have the disadvantage of hiding the raw data behind the summary deliverables. If you are conducting games user research, it is incumbent upon you to understand where your data are coming from and what they mean. Doing so enables you to do deeper analysis, as well as simply sanity checking that your results are meaningful and useful for your design partners.

A case study from our own research illustrates this challenge. We were tasked with determining how frequently users would look at certain locations in an Xbox app while carrying out a series of typical behaviors. The app team had two different conditions that interested them, which created a 2×2 factorial design; while we counterbalanced the conditions across participants accordingly, we were less focused on the interaction of the two conditions than on the main effects. For quantitative eye-tracking studies (the kind that produce the visually appealing heat maps), Nielsen Norman group recommend 30 participants per condition; in this instance, that would be 120 participants. In order to have a somewhat reasonable turnaround, we reduced this number to 90 and accepted that we would have reduced power in the interaction analysis (which, as mentioned earlier, was less critical). In the end, these 90 scheduled participants resulted in usable data from 72 participants,[2] enough for a reliable main effects analysis.

Because our research question was constrained, we were able to create a user scenario that could be completed in a 30-minute session (including calibration). We relied on overbooking the sessions to have a buffer of participants in case any one person was not compatible with the tracker, and running with multiple trackers in multiple rooms. Additionally, we were able to use support staff (two full-time employees, seven contract staff working in shifts, all trained on the Tobii software) to do immediate processing of the sessions as they came in—specifically, they defined the periods during the sessions when participants were on the app home screen, which was the segment of interest for this study. All of these factors allowed us to collect the data within one week, and process the data at the same time, significantly reducing the turnaround time. Having a very constrained set of research questions then allowed us to get a report back to the team in a couple of days.

[2] This is an unavoidable risk with some people and some trackers—it is just not possible to get good calibration and reliable data. More modern trackers tend to be more robust, and having proper environmental conditions (lighting, chairs that force people to sit upright, adjustable tables for the trackers) can help with this. But it is important to take into account that sometimes, some participants will just not provide good data.

We describe this to illustrate how a properly organized quantitative study requires extensive effort and resources. Before deciding whether to engage in this sort of data collection, it is critical to consider whether the information it will provide exceeds what you could get from more traditional methods enough to justify the costs. Here, knowing where people were looking in the app (and how much time they spent looking at different things) was central to what the team needed to know.

The challenge of interpreting data

All of which raises the question of *why* it is necessary to conduct studies with this number of participants. One answer has to do with simple reliability. Pernice & Nielsen (2009) describe the impact that low numbers of participants have on eye-tracking heat maps, and their interpretations. With these measures, running 6–10 participants makes it much more likely that a single outlier can skew the results in an inaccurate direction. On top of that, these data streams are typically noisy. One person can have dramatically different ranges for EDA or heart-rate variability than another (indeed, some participants show no such responses!). Collecting sufficient numbers is necessary for making the data interpretable.

What about usability and examining the data one person at a time in a small sample of players? Certainly, many of the studies and articles that have been written about biometrics in games have used that approach. The difficulty here is that often these demonstrations take a discount usability approach to what is fundamentally *opinion* data. Thinking back to playtests, it is clear why you would not ask one person (or two, or five) how fun they thought your game was and leave it at that. It is necessary to collect a statistically reliable sample so that you have a better estimate of what the broader audience will think of your title. This stands in contrast to traditional usability, where the failure of a single participant can point out a problem with your design.

The constructs that biometric data typically measure (engagement, enjoyment, frustration) are fundamentally opinion measures. They describe how the player "feels" about the game. For these measures to be meaningful, they must be able to estimate the population-level response. While this is relatively easy to accomplish in survey form (ask a large enough number of people questions after meaningful gameplay period or a set amount of time), aggregating biometric responses is much trickier because these measures are noisy and the volume of data over time is so much greater than survey data. Consider the following example: two people play a platformer (such as Super Meat Boy) for 5 minutes. The game has a set of short, well-defined levels that proceed in sequence, so there is little variability in potential gameplay (in contrast to a more open world game). Player 1 dies twice, Player 2 dies four times. Even though they have played the same game for the same amount of time, the two players have had very different experiences, and their biometric responses are likely to be quite out of sync. Thus, a simple linear averaging will not work.

One way to overcome this is to combine biometric data with design intelligence focused in-game telemetry. This is an approach we have examined in Studios User Research, and the method shows promise. To achieve this, it is first necessary to have the game software track important events and player actions as they occur in the game (as discussed in the Design Intelligence section). The telemetry system then needs to communicate with the biometric data recording system, so that the timestamps on all recordings are in sync (a Network Time Protocol can be helpful in this regard). Once the events and actions that characterize and individual player's experience are aligned with that player's biometric data, it is possible to aggregate the data from multiple players in a way that summarizes players' reactions overall.

An example illustrates the integration of player telemetry and biometrics that we achieved in our labs. These examples involve the collection of EDA (electrodermal activity; fluctuation in perspiration levels related to arousal) data using internally produced hardware from participants in playtest sessions. Figure 16.14 is taken from data collected during tests of a multiplayer mode in a game. We took the EDA data from all participants 30 s before and 30 s after a failure event in that mode, and averaged those data together. Here the EDA data is time locked to a single category of event. This makes the data at a given point in time roughly equivalent across participants; the +10 s point on the graph represents 10 s after failure for everyone. As can be seen in the graph, the EDA peaks approximately 5 s after the event. Given the known latency of EDA (1–5 s, depending on circumstances), this is consistent with an increase in arousal in response to the failure state. We have conducted similar experiments by mapping player EDA to location in the game world, to create images of (potentially) higher and lower arousal based on in-game position.

This combination of game telemetry and biometric data points to another form of analysis that can be useful: treating biometric data as *human telemetry*. Analysis algorithms can be used to identify time points of peak excitement or frustration for individual players. These time points can then be mapped onto moments in time during the game or locations in the game world, to identify points when players were more (or less) engaged. This method is analogous to the approach taken with biometric storyboards but done in aggregate with multiple participants. Such a method allows

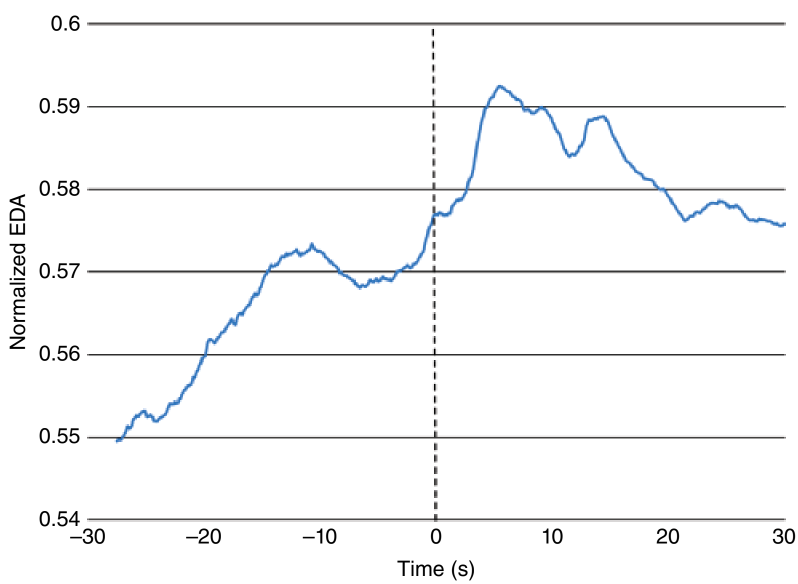


Figure 16.14 EDA averaged across multiple participants and multiple instances of a failure event in a multiplayer game mode. Time 0 (dashed line) represents the occurrence of the event. EDA is normalized to a 0–1 scale for all participants.

for potentially meaningful comparisons: are people as excited by the last boss as the first? Do we see a lot of players getting bored in the open-world areas? What are the response to the major plot twist?

The above description assumes the mapping of a single measure (e.g., EDA), with game data. However, having multiple measures (e.g., EDA with heart-rate variability and facial EMG) can provide convergent lines of evidence and make it easier to pinpoint users' emotional states. When possible, we recommend having convergent biometric data streams, or, at a minimum, biometric data coupled with behavioral or attitudinal data. As described in the Design Intelligence section, such converging data streams are required to have a truly holistic view of the player experience.

We have spent some time here arguing for the importance of aggregation in biometric data to draw meaningful conclusions. This does not necessarily mean that such methods should *never* be used in a small-N usability setting. For example, eye tracking can be considered an additional behavioral data stream during usability, indicating whether a participant failed to look at an important cue, or failed to understand. During testing of an early build of Quantum Break (2016), eye tracking was used in this manner to verify that several usability participants failed to understand the context of a particular scene because they were focused on the main character climbing up a ladder, rather than looking at what was happening in the background (see Figure 16.15). Similarly, EDA and heart-rate variability can indicate when participants are becoming especially frustrated during a certain scenario that they cannot complete (whether such frustration would or would not be otherwise apparent is left as a question for the experimenter).

Figure 16.15 A usability participant during testing of an early build of Quantum Break (2016). Eye tracking (red dot and line overlaid on game screen) show that the participant was focused on climbing up the ladder, rather than the plot-critical ship in the background (left side of screen).

The issue that the researcher must always consider is *what is the value?* What questions are you answering, and could they be answered in another (less costly, easier to analyze) way? For many teams, it may be the case that biometrics do not offer sufficient information to justify the cost in time and money that it takes to implement procedures, collect data, and perform analysis. These methods can be flashy and technically impressive, but if they cannot be applied to improve the design of the game then they have accomplished nothing. As has been demonstrated earlier in the chapter, there is much that can be learned from simply observing people play your game, and even more that can be gathered from simple surveys. For most groups, these straightforward methods will have the biggest impact by far. That said, as equipment costs go down, and as more groups explore these techniques, the resources available to even smaller teams will grow. Over the next several years, it could be relatively cheap to outfit a lab with reliable biometric sensors and open-source analysis packages could make data interpretation more straightforward. The challenge will lay once again with the imagination of the UR to determine what they can learn and how they can help their partners.

Lessons learned

Having experimented with various forms of biometric measurement in various settings, we propose the following question to any groups considering employing these methods: Do you really need this data to help make your product better? We also propose the following answer: Probably not. Unlike the other methods discussed in this chapter, biometrics data is inessential. There are certain very specific questions for which biometrics can provide direct answers (especially eye tracking), but aside from that the information gathered may not be worth the time and effort, especially when many less costly methods can have such a profound impact.

If you are going to pursue biometrics, it is critical to have clear, easily answered questions in mind. If you can do direct comparisons between conditions, that is an ideal scenario. Barring that, at least have a clear idea of what parts of the experience you will focus on, and come up with some sort of a priori hypothesis of what you expect to see. If you go into this form of analysis thinking “Let’s see what happened—surely there will be patterns in the data!” you will lose hours exploring every possible trend and interaction, and very likely not come up with a clear-cut recommendation. Remember, the goal with all of this is to provide your team with information that can help them make their product better—spinning your wheels on analysis and noticing a minor trend in a subsection of one level probably isn’t going to do that.

It is also important to give your participants advance warning of what you will be doing during the session. Not everyone will be comfortable with some of the biometric data collection methods you are using (or even with the idea of that data being collected), so it is important to give them the choice to opt out both when they are recruited *and* when they come in for the session. Have a clear, concise description of what data you are collecting and how it will be collected on hand to give to the participant, and update any NDAs or consent forms accordingly. The participants are being gracious enough to come in and test your game for you, so it’s important to respect their wishes.

Be aware, too, that these data need work to interpret and analyze. Very rarely will you be able to use the raw data straight away. Some software programs will do this for you, but even then you may have to make decisions about what is being done to your data (e.g., what sort of filter is applied, what is your temporal window, do you care about tonic or phasic signals). Again, it's best to know the answers to all of this before you've collected a single data point, so that you have your approach clear in your mind.

Finally, in the leadup to a study, make sure all of your equipment is in order and that you are collecting clear and interpretable data. Then check it again. Then do it one more time. An example: when using prerecorded videos for eye tracking, having everything in the proper format is critical. We have lost data because an oddity of the video resolution introduced an error into the tracking results, and we only saw it when examining postsession visualizations. You are going to have many systems working together, so make sure that all of these systems are properly communicating with each other. Do not trust that anything will "just work," because it won't. If you use some disposable materials, such as prepackaged electrodes, be sure you have enough on hand (we've had to depend on the generosity of nearby labs more than once). When working with biometric data sources, process is key. Things will go wrong, equipment will break, and setting up will take longer than expected. The best practice is to have a clearly established process in place for every step of the study so that you know when you need things, how things need to be set up, how the data will be collected and delivered, what you are going to do with it, and you have buffer built into every step because something always goes wrong.

Conclusion

Applied games user research has a long history of leveraging social science research methods to improve the player experience with video games. At Microsoft Studios User Research, our methods have evolved since the late 1990s, and we look forward to the innovation in the industry in future years. In this chapter, we've presented a number of different techniques that we have developed to ensure that the player's experience with our games is the best it possibly can be, and matches the designer's intent. For us to be able to accomplish what we have, we've built on the foundation that others inside and outside the field have provided. We hope that our additions to this foundation provide a jumping-off point for games user researchers for generations to come, and we look forward to helping build that future with others in the field.

There's one last *lesson learned* to think about when applying the collective knowledge of our discipline to your products: as games user researchers, it can be easy to get excited about the shiny new toys of research methods and techniques but being effective with any of them requires a strong focus on the fundamentals. You are, fundamentally, the unbiased feedback loop from the player to the designer, and that's only possible with a focus on the three keys of games user research: researcher, method, and outcome. Continually keep your research goals and questions in mind. Always apply the right method at the right time to answer your research questions. And lastly, always provide the development team with answers to those questions in the form of insight that they can act on. It is then, and only then, that the true voice of the player can be heard.

References

Abbott, P. H. (2002). *The Cambridge introduction to narrative.* Cambridge: Cambridge University Press.

Amaya, G., Davis, J., Gunn, D., Harrison, C., Pagulayan, R., Phillips, B., & Wixon, D. (2008). Games user research (GUR): Our experience with and evolution of four methods. In K. Isbister & N. Schaffer, (Eds.), *Game usability: Advice from the experts for advancing the player experience* (pp. 35–64). Burlington, MA: Morgan Kaufmann Publishers.

Benveniste, E. (1971). *Problems in general linguistics* (M. E. Meet, Trans.). Coral Gables, FL: University of Miami Press.

Chatman, S. (1978). *Story and discourse: Narrative structure in fiction and film.* Ithaca, NY: Cornell University Press.

D4: Dark Dreams Don't Die (2014). [Computer software]. Redmond, WA: Microsoft Studios

Fairclough, S. (2009). Fundamentals of physiological computing. *Interacting with Computers*, 21(1–2), 133–145.

Figner, B., & Murphy, R. O. (2010). Using skin conductance in judgment and decision making research. In M. Schulte-Mecklenbeck, A. Kuehberger, & R. Ranyard (Eds.), *A handbook of process tracing methods for decision research.* New York, NY: Psychology Press.

Fulton, B. (2002). Beyond psychological theory: Getting data that improve games. *Proceedings of the Game Developer's Conference,* San Jose, CA. Retrieved from http://www.gamasutra.com/view/feature/131412/beyond_psychological_theory_.php

Genette, G. (1988). *Narrative discourse revisited* (J. E. Lewin, Trans.). Ithaca, NY: Cornell University Press.

Gran Turismo (1997). [Computer software]. Tokyo, Japan: Sony Computer Entertainment.

Hagen, J. (2015, August). *Communicating about UR to management.* Paper presented at the 2015 Bungie User Research Summer Camp, Bellevue, WA.

Halo 4 (2012). [Computer software]. Redmond, WA: Microsoft Studios.

Hendersen, D. (2014a). *Using user research to improve game narratives.* Paper presented at the 2014 Game Developers Conference, San Francisco, CA. Retrieved from http://www.gdcvault.com/play/1020031/Using-User-Research-to-Improve

Hendersen, D. (2014b). *Narrative usability.* Paper presented at the 2014 Games User Research Summit, San Francisco, CA.

Herbst, S. & Medlock, M. (2014, April). It's a trap! Three remarkably common UI traps and how to avoid them in your designs [Video file]. Presented at the 2014 "Build" Conference, San Francisco, CA. Retrieved from https://channel9.msdn.com/Events/Build/2014/2-660

Hopson, J. (2006, November 10). We're not listening: An open letter to academic game researchers. *Gamasutra,* 6. Retrieved from http://www.gamasutra.com/view/feature/1783/were_not_listening_an_open_.php

Kim, J., Gunn, D., Schuh, E., Phillips, B., Pagulayan, R., & Wixon, D. (2008). Tracking Real-Time User Experience (TRUE): A comprehensive instrumentation solution for complex systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Florence, Italy) (pp. 443–452). New York, NY: ACM.

Leedy, P. L., & Ormrod, J. E. (2004). *Practical research: Planning and design* (8th ed.). Upper Saddle River, NJ: Prentice Hall.

McCabe, D. P., & Castel, A. D. (2008). Seeing is believing: The effect of brain images on judgments of scientific reasoning. *Cognition*, 107(1), 343–352.

Medlock, M. C. (2014, September). *History of video games user research.* Paper presented at the 2014 Games User Research Summit, San Francisco, CA.

Mirza-Babaei, P., & McAllister, G. (2011, May). *Biometric storyboards: Visualizing meaningful gameplay events.* Paper presented at CHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada

Nielsen, J. (1993). *Usability engineering*. San Francisco, CA: Morgan Kaufmann.
Olson, C. K. (2010). Children's motivations for video game play in the context of normal development. *Review of General Psychology*, 14(2), 180–187.
Pagulayan, R. J., Gunn, D. V., & Romero, R. L. (2006). A gameplay-centered design framework for human factors in games. In W. Karwowski (Ed.), *International encyclopedia of ergonomics and human factors* (2nd ed.). London: Taylor & Francis.
Pagulayan, R. J., Keeker, K., Fuller, T., Wixon, D., Romero, R. L., & Gunn, D. V. (2012). User-centered design in games. In J. A. Jacko, *Human–computer interaction handbook: Fundamentals, evolving technologies, and emerging applications* (3rd ed.). Boca Raton, FL: CRC Press.
Pagulayan, R. J., Steury, K. R., Fulton, B., & Romero, R. L. (2003). Designing for fun: Usertesting case studies. In M. Blythe, K. Overbeeke, A. Monk, & P. Wright (Eds.), *Funology: From usability to enjoyment*. New York, NY: Springer.
Pernice, K., & Nielsen, J. (2009). Eyetracking Methodology: How to Conduct and Evaluate Usability Studies Using Eyetracking [Guide]. Downloaded from https://www.nngroup.com/reports/how-to-conduct-eyetracking-studies/
Pokémon: Red and Green (1996). [Computer software]. Kyoto, Japan: Nintendo.
Quantum Break (2016). [Computer software]. Redmond, WA: Microsoft Studios.
Ryse: Son of Rome (2013). [Computer software]. Redmond, WA: Microsoft Studios.
Sasse, D. B. (2008) A framework for psychophysiological data acquisition in digital games (Unpublished dissertation). Otto-von-Guericke Universität, Magdeburg, Germany.
Schuh, E., Gunn, D., Phillips, B., Pagulayan, R., Kim, J., & Wixon, D. (2008). TRUE instrumentation: Tracking real-time user experience in games. In K. Isbister, & N. Schaffer, *Game usability: Advice from the experts for advancing the player experience* (pp. 237–265). Burlington, MA: Morgan Kaufmann Publishers.
Skinner, B. F. (1932). On the rate of formation of a conditioned reflex. *Journal of General Psychology*, 7, 274–286.
Skinner, B. F. (1938). *The behavior of organisms: An experimental analysis*. New York, NY: Appleton-Century.
Stanovich, K. (2007). *How to think straight about psychology* (8th ed.). Boston, MA: Allyn & Bacon.
Thompson, C. (2007, September). Halo 3: How Microsoft labs invented a new science of play. *Wired*, 15, 140–147.
Voodoo Vince (2003). [Computer software]. Redmond, WA: Microsoft.
Wickens, C. D., & Hollands, J. G. (2000). *Engineering psychology and human performance* (3rd ed.). Upper Saddle River, NJ: Prentice Hall.

Part V
Input / Output

17

Fitts' Law

I. Scott MacKenzie

Introduction

Human movement is ubiquitous in computing. Our arms, wrists, and fingers busy themselves on keyboards, desktops, and contact-sensitive displays. So, matching the movement limits and capabilities of humans with interaction techniques on computing systems is an important area of research in human-computer interaction (HCI). Considerable HCI research is directed at modeling, predicting, and measuring human performance. In the realm of human movement, Fitts' law is the preeminent model for this research.

The full spectrum of human movement applicable to Fitts' law is broader than the three examples—arms, wrists, and fingers—given in the preceding paragraph. In contexts such as gaming, virtual reality, or accessible computing, movements may also involve the torso, legs, feet, eyes, face, tongue, lip, skin, head, and so on. Notably, for each of these input modalities, there are examples where Fitts' law was used to explore the design space or to quantify human performance.

This chapter provides an overview of Fitts' law. As we shall see, Fitts' law is a model both for predicting and measuring. For predicting, Fitts' law is an equation giving the time to acquire and select a target based on the distance moved and the size of the target. For measuring, Fitts' law provides a method to quantify human performance in a single measure, "throughput." Throughput, when calculated as described later in this chapter, combines speed and accuracy in performing a target acquisition task.

We begin with background details and a brief tour of Fitts' law, and follow by describing refinements to correct flaws or to improve the model's prediction power or theoretical basis. Fitts' law evaluations of computer input techniques are more consistent in recent years due to the emergence of ISO 9241-9 (ISO, 2000), an ISO standard for evaluating input devices. The Fitts' law methods used in the standard are summarized and software tools are presented that implement the methods. Since Fitts' throughput is the main performance measure for such evaluations, we also detail the calculation of throughput according to best practice methods. We then present an example of the use of Fitts' law and ISO 9241-9 for measuring human performance. The example involves touch-based target selection on a mobile phone with a contact-sensitive display.

The Wiley Handbook of Human Computer Interaction, Volume 1, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

Background

Like many psychologists in the 1950s, Fitts was motivated to investigate whether human performance could be quantified using a metaphor from the new and exciting field of information theory. This field emerged from the work of Shannon, Wiener, and other mathematicians in the 1940s. The terms *probability*, *redundancy*, *bits*, *noise*, and *channels* entered the vocabulary of experimental psychologists as they explored the latest technique of measuring and modeling human behavior. Two well-known models in this vein are the Hick–Hyman law for choice reaction time (Hick, 1952; Hyman, 1953) and Fitts' law for the information capacity of the human motor system (Fitts, 1954).

Fitts' particular interest was rapid-aimed movements, where a human operator acquires or selects targets of a certain size over a certain distance. Fitts proposed a model—now "law"—that is widely used in fields such as ergonomics, engineering, psychology, and human-computer interaction. The starting point for Fitts' law is an equation known as Shannon's *Theorem 17*, which gives the information capacity C (in bits/s) of a communications channel of bandwidth B (in s^{-1} or Hz) as

$$C = B \log_2\left(\frac{S}{N}+1\right) \tag{17.1}$$

where S is the signal power and N is the noise power (Shannon & Weaver, 1949, pp. 100–103). Fitts reasoned that a human operator that performs a movement over a certain amplitude to acquire a target of a certain width is demonstrating a "rate of information transfer" (Fitts, 1954, p. 381). In Fitts' analogy, movement amplitudes are like signals and target tolerances or widths are like noise.

Fitts proposed an *index of difficulty* (ID) for a target acquisition task using a log term slightly rearranged from Eq. 17.1. Signal power (S) and noise power (N) are replaced by movement amplitude (A) and target width (W), respectively:

$$ID = \log_2\left(\frac{2A}{W}\right) \tag{17.2}$$

Fitts referred to the target width as the "permissible variability" or the "movement tolerance" (Fitts, 1954, p. 382). This is the region within which a movement is terminated. As with the log term in Eq. 17.1, the units for ID are *bits* because the ratio within the parentheses is unitless and the log is taken to base 2.

Fitts' idea was novel for two reasons: First, it suggested that the difficulty of a target selection task could be quantified using the information metric *bits*. Second, it introduced the idea that the act of performing a target selection task is akin to transmitting information through a channel—a human channel. Fitts called the rate of transmission the *index of performance*, although today the term *throughput* (TP) is more common. (For consistency, the term *throughput* is used throughout this chapter.)

Throughput is calculated over a sequence of trials as a simple quotient. The index of difficulty (ID) of the task is the numerator and the mean movement time (MT) is the denominator:

$$TP = \frac{ID}{MT} \tag{17.3}$$

With *ID* in bits and *MT* in seconds, *TP* has units *bits per second* or *bits/s*. A central thesis in Fitts' work is that throughput is independent of movement amplitude and target width, as embedded in *ID*. In other words, as *ID* changes (due to changes in *A* or *W*), *MT* changes in an opposing manner and *TP* remains more-or-less constant.

Of course, throughput is expected to be influenced by other factors, such as device, interaction property, or environment. Two devices were compared in Fitts' original experiment (see next section). In HCI, a myriad of factors, or independent variables, can be explored using Fitts' throughput as a dependent variable. Examples include "device" (mouse versus stylus versus trackball—see MacKenzie, Sellen, & Buxton, 1991), "dwell interval" with an eye tracker (700 ms vs. 500 ms—see Zhang & MacKenzie, 2007), or "device position" (supported versus mobile—see MacKenzie, 2015). Throughput is particularly appealing as a dependent variable because it combines speed and accuracy in a single measure (using a technique described shortly).

Of the two uses of Fitts' law noted above—predicting and measuring—throughput exemplifies the use of Fitts' law for measuring.

Fitts' Experiments

The original investigation (Fitts, 1954) involved four experiment conditions: two reciprocal or serial tapping tasks (1 oz stylus and 1 lb stylus), a disc transfer task, and a pin transfer task. For the tapping condition, a participant moved a stylus back and forth between two plates as quickly as possible and tapped the plates at their centers (see Figure 17.1a). Fitts later devised a discrete variation of the task (Fitts & Peterson, 1964). For the discrete task, the participant selected one of two targets in response to a stimulus light (see Figure 17.1b). The tasks in Figure 17.1 are commonly called the "Fitts' paradigm." It is easy to imagine how to update Fitts' apparatus using contemporary computing technology.

Fitts published summary data for his 1954 experiments, so a reexamination of his results is possible. For the stylus-tapping conditions, four target amplitudes (*A*) were crossed with four target widths (*W*). For each *A*-*W* condition, participants performed

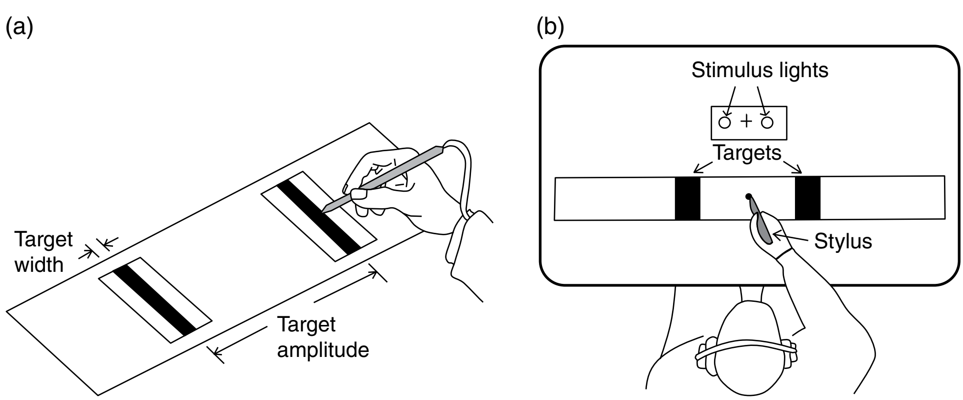


Figure 17.1 The Fitts paradigm. (a) serial tapping task (after Fitts, 1954) (b) discrete task (after Fitts & Peterson, 1964).

Table 17.1 Data from Fitts' (1954) serial tapping task experiment with a 1 oz stylus. An extra column shows the effective target width (W_e) after adjusting W for the percentage errors (see text).

A (in)	*W (in)*	*W_e (in)*	*ER (%)*	*ID (bits)*	*MT (ms)*	*TP (bits/s)*
2	0.25	0.243	3.35	4	392	10.20
2	0.50	0.444	1.99	3	281	10.68
2	1.00	0.725	0.44	2	212	9.43
2	2.00	1.020	0.00	1	180	5.56
4	0.25	0.244	3.41	5	484	10.33
4	0.50	0.468	2.72	4	372	10.75
4	1.00	0.812	1.09	3	260	11.54
4	2.00	1.233	0.08	2	203	9.85
8	0.25	0.235	2.78	6	580	10.34
8	0.50	0.446	2.05	5	469	10.66
8	1.00	0.914	2.38	4	357	11.20
8	2.00	1.576	0.87	3	279	10.75
16	0.25	0.247	3.65	7	731	9.58
16	0.50	0.468	2.73	6	595	10.08
16	1.00	0.832	1.30	5	481	10.40
16	2.00	1.519	0.65	4	388	10.31
				Mean	391.5	10.10
				SD	157.3	1.33

two sequences of trials lasting 15 s each. (In current practice, a "sequence" is usually a specified number of trials, for instance 25, rather than a specified time interval.) The summary data for the 1 oz stylus condition are given in Table 17.1. As well as *A* and *W*, the table includes the error rate (*ER*), index of difficulty (*ID*), movement time (*MT*), and throughput (*TP*). The effective target width (W_e) column was added, as discussed shortly.

The combination of conditions in Table 17.1 yields task difficulties ranging from 1 bit to 7 bits. The mean *MTs* observed ranged from 180 ms ($ID = 1$ bit) to 731 ms ($ID = 7$ bits), with each mean derived from more than 600 observations over 16 participants. The standard deviation in the *MT* values was 157.3 ms, which is 40.2% of the mean. This is fully expected since "hard tasks" (e.g., $ID = 7$ bits) will obviously take longer than "easy tasks" (e.g., $ID = 1$ bit).

Fitts calculated throughput by dividing *ID* by *MT* (Eq. 17.3) for each task condition. The mean throughput was 10.10 bits/s. A quick glance at the *TP* column in Table 17.1 shows strong evidence for the thesis that the rate of information processing is relatively independent of task difficulty. Despite the wide range of task difficulties, the standard deviation of the *TP* values was 1.33 bits/s, which is just 13.2% of the mean.

One way to visualize the data in Table 17.1 and the independence of *ID* on *TP* is through a scatter plot showing the *MT-ID* point for each task condition. Figure 17.2 shows such a plot for the data in Table 17.1. The figure also includes the best fitting line (via least-squares regression), the linear equation, and the squared correlation. The independence of *ID* on *TP* is reflected in the closeness of the points to the regression line (indicating a constant *ID/MT* ratio). Indeed, the fit is very good with 96.6% of the variance explained by the model.

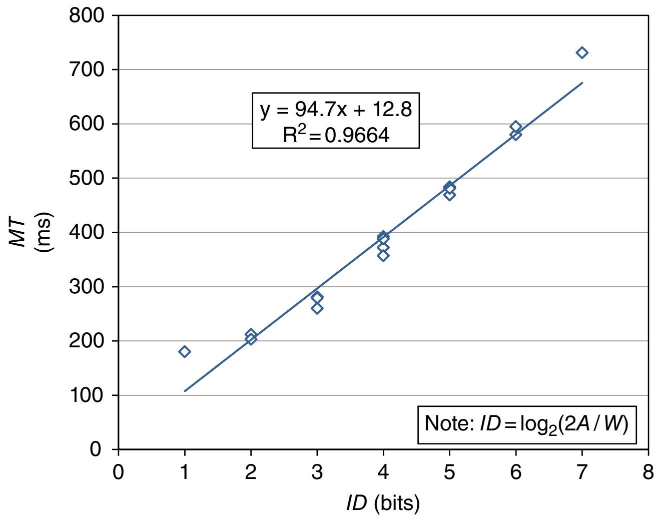


Figure 17.2 Scatter plot and least-squares regression analysis for the data in Table 17.1.

The linear equation in Figure 17.2 takes the following general form:

$$MT = a + b\,ID \tag{17.4}$$

The regression coefficients include an intercept *a* with units *seconds* and a slope *b* with units *seconds per bit*. (Several interesting yet difficult issues arise in interpreting the slope and intercept coefficients in Eq. 17.4. Due to space limitations, these are not elaborated here. The interested reader is directed to sections 3.4 and 3.5 in Soukoreff & MacKenzie, 2004.) Equation 17.4 exemplifies the use of Fitts' law for *predicting*. This is in contrast with Eq. 17.3 which is the use of Fitts' law for *measuring*.

Refinements to Fitts' Law

In the years since the first publication in 1954, many changes or refinements to Fitts' law have been proposed. While there are considerations in both theory and practice, a prevailing rationale is the need for precise mathematical formulations in HCI and other fields for the purpose of measurement. One can imagine (and hope!) that different researchers using Fitts' law to examine similar phenomena should obtain similar results. This is only possible if there is general agreement on the methods for gathering and applying data.

An early motivation for altering or improving Fitts' law stemmed from the observation that the *MT-ID* data points curved away from the regression line, with the most deviate point at $ID = 1$ bit. This is clearly seen in the leftmost point in Figure 17.2. In an effort to improve the data-to-model fit, Welford (1960, 1968, p. 147) introduced the following formulation:

$$ID = \log_2\left(\frac{A}{W} + 0.5\right) \tag{17.5}$$

This version of *ID* has been used frequently over the years, and in particular by Card, English, and Burr (1978) in their comparative evaluation of the computer mouse. (A reanalysis of the results reported by Card et al., 1978, is given by MacKenzie and Soukoreff, 2003, in view of a contemporary understanding of Fitts' law.) Fitts also used the Welford formulation in a 1968 paper and reported an improvement in the regression line fit compared to the Fitts formulation (Fitts & Peterson, 1964, p. 110).

In 1989, it was shown that Fitts deduced his relationship citing an approximation of Shannon's theorem that only applies if the signal-to-noise ratio is large (Fitts, 1954, p. 388; Goldman, 1953, p. 157; MacKenzie, 1989, 1992). The signal-to-noise ratio in Shannon's theorem appears as the *A*-to-*W* ratio in Fitts' analogy. As seen in Table 17.1, the *A*-to-*W* ratio in Fitts' stylus-tapping experiment extended as low as 1 : 1! The variation of Fitts' index of difficulty suggested by direct analogy with Shannon's information theorem is

$$ID = \log_2\left(\frac{A}{W}+1\right) \tag{17.6}$$

Besides the improved link with information theory, Eq. 17.6, known as the *Shannon formulation*, provides better correlations compared to the Fitts or Shannon formulation (MacKenzie, 1989, Table 1 and Table 2; 1991, Table 4; 2013, Table 3).

An additional feature of the Shannon formulation is that *ID* cannot be negative. Obviously, a negative rating for task difficulty presents a serious theoretical problem. Although the prospect of a negative *ID* may seem unlikely, such conditions have actually been reported in the Fitts' law literature (Card et al., 1978; Crossman & Goodeve, 1983; Gillan, Holden, Adam, Rudisill, & Magee, 1990; Ware & Mikaelian, 1987). With the Shannon formulation, a negative *ID* is simply not possible. This is illustrated in Figure 17.3, which shows *ID* smoothly approaching 0 bits as *A* approaches 0. With the Fitts and Welford formulations, *ID* dips negative for small *A*.

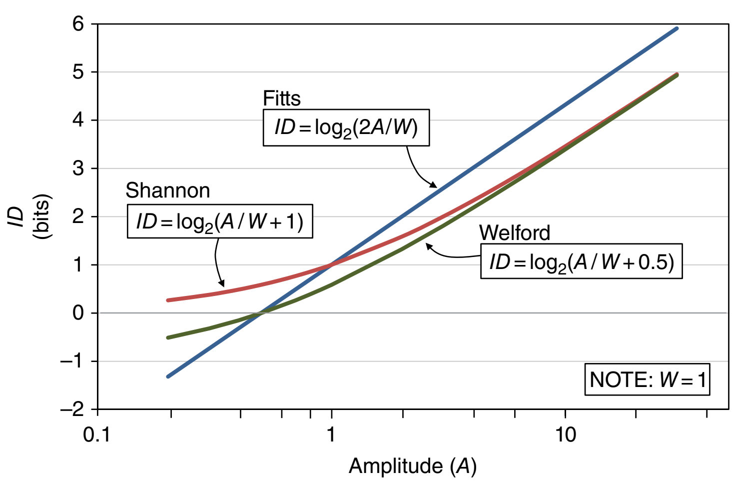


Figure 17.3 With the Shannon formulation, *ID* approaches 0 bits as *A* approaches 0.

Adjustment for Accuracy

Of greater practical importance is a technique to improve the information-theoretic analogy in Fitts' law by adjusting the specified or set target width (akin to noise) according to the spatial variability in the human operator's output responses over a sequence of trials. The idea was first proposed by Crossman in 1957 in an unpublished report (cited in Welford, 1968, p. 147). Use of the adjustment was later examined and endorsed by Fitts (Fitts & Peterson, 1964, p. 110).

The output or *effective target width* (W_e) is derived from the distribution of "hits" (see MacKenzie, 1992, section 3.4; Welford, 1968, pp. 147–148). This adjustment lies at the very heart of the information-theoretic metaphor—that movement amplitudes are analogous to "signals" and that endpoint variability (viz., target width) is analogous to "noise." In fact, the information theorem underlying Fitts' law assumes that the signal is "perturbed by white thermal noise" (Shannon & Weaver, 1949, p. 100). The analogous requirement in motor tasks is a Gaussian or normal distribution of hits—a property observed by numerous researchers (e.g., Fitts & Radford, 1966; MacKenzie, 1991, p. 84; 2015; Welford, 1968, p. 154).

The experimental implication of normalizing output measures is illustrated as follows. The entropy, or information, in a normal distribution is $\log_2\left(\sqrt{2\pi e}\,\sigma\right) = \log_2(4.133\,\sigma)$, where σ is the standard deviation in the unit of measurement. Splitting the constant 4.133 into a pair of *z*-scores for the unit-normal curve (i.e., $\sigma = 1$), one finds that 96% of the total area is bounded by $-2.066 < z < +2.066$. In other words, a condition that target width is analogous to noise is that the distribution is normal with 96% of the hits falling within the target and 4% of the hits missing the target. See Figure 17.4a. When an error rate other than 4% is observed, target width is adjusted to form the effective target width in keeping with the underlying theory.

There are two methods for determining the effective target width, the *standard-deviation method* and the *discrete-error method*. If the standard deviation of the endpoint coordinates is known, just multiply *SD* by 4.133 to get W_e. If only the percentage of errors is known, the method uses a table of *z*-scores for areas under the unit-normal curve. (Such a table is found in the appendices of most statistics textbooks; *z*-scores are also available using the NORM.S.INV function in Microsoft Excel.) Here is the method: if *n* percent errors are observed over a sequence of trials for a particular *A*-*W* condition, determine *z* such that $\pm z$ contains $100 - n$ percent of the area under the unit-normal curve. Multiply *W* by $\frac{2.066}{z}$ to get W_e. As an example, if 2% errors occur on a sequence of trials when selecting a 5 cm wide target, then $W_e = \frac{2.066}{z} \times W = \frac{2.066}{2.326} \times 5 = 4.45\,\text{cm}$. See Figure 17.4b. Broadly, the figure illustrates that $W_e < W$ when error rates are less than 4% and that $W_e > W$ when error rates exceed 4%.

Experiments using the adjusted or effective target width will typically find a reduced variation in *IP* because of the speed-accuracy tradeoff: participants who take longer are more accurate and demonstrate less endpoint variability. Reduced endpoint variability decreases the effective target width and therefore increases the effective index of difficulty (see Eq. 17.3). The converse is also true. On the whole, an increase

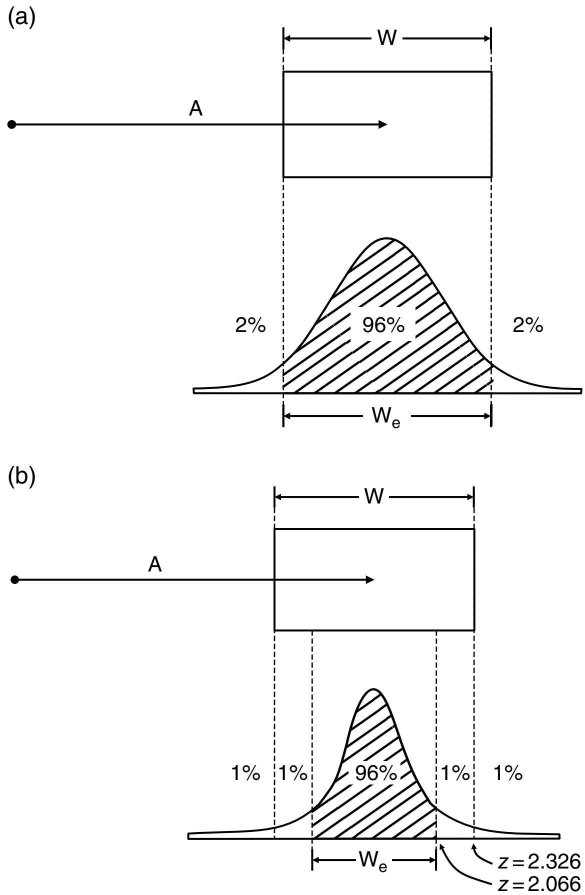


Figure 17.4 Method of adjusting target width based on the distribution of selections. (a) When 4% errors occur, the effective target width, $W_e = W$. (b) When less than 4% errors occurs, $W_e < W$.

(or decrease) in *MT* is accompanied by an increase (or decrease) in the effective *ID*, and this tends to lessen the variability in *IP* (see Eq. 17.2).

The technique just described dates to 1957, yet it was largely ignored in the published body of Fitts' law research that followed.[1] There are several possible reasons. First, the method is tricky and its derivation from information-theoretic principles is complicated (see Reza, 1961, pp. 278–282). Second, selection coordinates must be recorded for each trial in order to calculate W_e from the standard deviation. This is feasible using a computer for data acquisition and statistical software for analysis, but manual measurement and data entry are extremely awkward.

[1] Since the early 1990s, use of the effective target width has increased, particularly in human-computer interaction. This is in part due to the recommended use of W_e in the performance evaluations described in ISO 9241-9 (ISO, 2000). The first use of W_e in HCI is the Fitts' law study described by MacKenzie, Sellen, and Buxton (1991).

Inaccuracy may enter when adjustments use the percent errors—the discrete-error method—because the extreme tails of the unit-normal distribution are involved. It is necessary to use *z*-scores with at least three decimal places of accuracy for the factoring ratio (which is multiplied by W to yield W_e). Manual lookup methods are prone to precision errors. Furthermore, some of the easier experimental conditions may have error rates too low to reveal the true distribution of hits. The technique cannot accommodate "perfect performance"! An example appears in Table 17.1 for the condition $A = W = 2$ in. Fitts reported an error rate of 0.00%, which seems reasonable because the target edges were touching. This observation implies a large adjustment because the distribution is very narrow in comparison to the target width over which the hits should have been distributed—with 4% errors! A pragmatic approach in this case is to assume a worst-case error rate of 0.0049% (which rounds to 0.00%) and proceed to make the adjustment.

Introducing a post hoc adjustment on target width as just described is important to maintain the information-theoretic analogy. There is a tacit assumption in Fitts' law that participants, although instructed to move "as quickly and accurately as possible," balance the demands of tasks to meet the spatial constraint that 96% of the hits fall within the target. When this condition is not met, the adjustment should be introduced. Note as well that if participants slow down and place undue emphasis on accuracy, the task changes; the constraints become temporal, and the prediction power of the model falls off (Meyer, Abrams, Kornblum, Wright, & Smith, 1988). In summary, Fitts' law is a model for rapid, aimed movements, and the presence of a nominal yet consistent error rate in participants' behavior is assumed, and arguably vital.

Table 17.1 includes an additional column for the effective target width (W_e), computed using the discrete-error method. A reanalysis of the data in Table 17.1 using W_e and the Shannon formulation for the index of difficulty is shown in Figure 17.5. The fit of the model is improved ($R^2 = .9877$) as the data points are now closer to the best fitting line. The curving away from the regression line for easy tasks appears corrected. Note that the range of *IDs* is narrower using adjusted

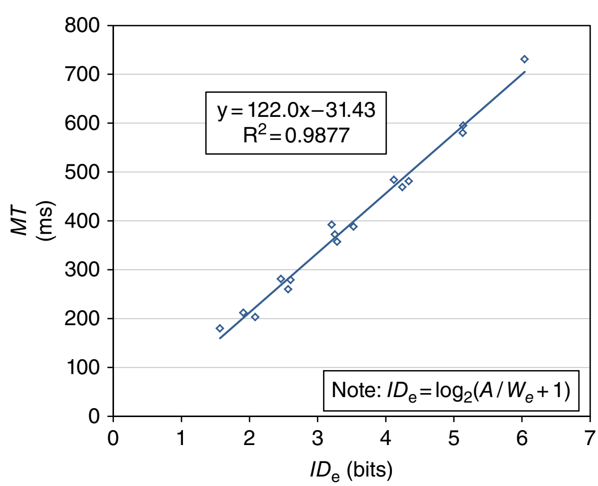


Figure 17.5 Reanalysis of data in Table 17.1 using the effective target width (W_e) and the Shannon formulation of index of difficulty (ID_e).

measures (cf. Figure 17.2 and Figure 17.5). This is due to the 1-bit decrease when *ID* is greater than about 2 bits (see Figure 17.3) and the general increase in *ID* for "easy" tasks because of the narrow distribution of hits.

Although Fitts' apparatus only recorded "hit" or "miss," modern computer-based systems are usually capable of recording the coordinate of target selection. (There are exceptions. Interaction methods that employ *dwell-time selection* perform target selection by maintaining the cursor within the target for a prescribed time interval. There is no selection coordinate per se. Examples of dwell-time selection include input using an eye tracker, such as MacKenzie, 2012, and Zhang & MacKenzie, 2007, or tilt-based input, such as Constantin & MacKenzie, 2014, and MacKenzie & Teather, 2012.)

As noted earlier, these data allow use of the standard-deviation method to calculate W_e. It is also possible, therefore, to calculate an effective amplitude (A_e)—the actual distance moved. The use of the A_e has little influence provided selections are distributed about the center of the targets. However, it is important to use A_e to prevent "gaming the system." For example, if all movements fall short and only traverse, say, $\frac{3}{4} \times A$, ID_e is artificially inflated if calculated using A. Using A_e prevents this. This is part of the overall premise in using "effective" values. Participants receive credit for what they actually did, not for what they were asked to do.

What is Fitts' Law?

At this juncture, it is worth stepping back and considering the big picture. What is Fitts' law? Among the refinements to Fitts' index of difficulty noted earlier, only the Welford and Shannon formulations were presented. Although other formulations exist, they are not reviewed here. There is a reason. In most cases, alternative formulations were introduced following a straightforward process: a change was proposed and rationalized and then a new prediction equation was presented and empirically tested for goodness of fit. Researchers often approach this exercise in a rather single-minded way. The goal is to improve the fit. A higher correlation is deemed evidence that the change improves the model—period. But there is a problem. The equation for the altered model often lacks any term with units "bits." So the information metaphor is lost. This can occur for a variety of reasons, such as using a nonlog form of *ID* (e.g., power, linear), inserting new terms, or splitting the log term into separate terms for *A* and *W*. If there is no term with units "bits," there is no throughput. While such models may indeed be valid, characterizing them as improvements to Fitts' law, or even as variations on Fitts' law is, arguably, wrong. They are entirely different models.

The position taken in the above paragraph follows from two points. First, the prediction form of Fitts' law (Eq. 17.4) does not appear in Fitts' original 1954 publication. Thus, it is questionable whether any effort motivated simply to improve the fit of the prediction equation falls within the realm of Fitts' law research. Second, Fitts' law is fundamentally about the *information capacity of the human motor system* (the title of Fitts' 1954 paper begins with the words set in italics). The true embodiment of Fitts' law is Eq. 17.3 for throughput, which appears in the original paper, albeit with different labels (Fitts, 1954, Eq. 2). Thus, retaining the information metaphor is central to Fitts' law.

ISO 9241-9

In the decades after the first publication (Fitts, 1954), numerous Fitts' law studies appeared—and in a great variety of forms. The internal validity of these studies is not in question, but there is considerable inconsistency in this body of research, and this renders across-study comparisons a daunting task. Simply put, it is often not possible to compare throughput values from one study to another. Reading carefully, details are often inadequately given. Where details are given, it is clear that throughput was often calculated in different ways. Furthermore, inconsistencies exist in the data collected or in the way the data are put to work in building Fitts' law models or calculating throughput. Clearly, Fitts' law research could benefit from a standardized methodology. This is particularly true in HCI, where the practical benefits of new ideas must be assessed and compared with related ideas in other publications. Enter ISO 9241-9.

ISO standards are written by technical committees drawn from the research and applied research communities. One standard relevant to HCI is the multipart ISO 9241, "Ergonomic requirements for office work with visual display terminals (VDTs)." Draft versions began to appear in the 1990s. Part 9 is "Requirements for non-keyboard input devices" (International Organization for Standardization, 2000). The standard has since been updated to the more generic title "Ergonomics of human-system interaction." The parts have also been updated, renamed, and renumbered. Part 9 is now Part 411, "Evaluation methods for the design of physical input devices" (International Organization for Standardization, 2012). (References in this chapter to ISO 9241-9 also apply to ISO 9241-411. With respect to the Fitts' law testing procedures, the two versions are the same.) The standard is relevant to virtually any input mechanism that can perform point-select operations on a computer. If there is one key benefit of ISO 9241-9, it is the standardization brought to the application of Fitts' law to input research in HCI.

The two main performance testing procedures in ISO 9241-9 employ the Fitts' paradigm. There is a one-dimensional (1D) task and a two-dimensional (2D) task, both using serial target selections. Including a 2D task is a pragmatic extension to Fitts' law to support interactions commonly found in computing systems. Although the possibility of a discrete task was described by Fitts (see Figure 17.1b) and is used in some Fitts' law studies, discrete tasks are not included in ISO 9241-9.

Screen snaps from the author's implementations are shown in Figure 17.6a for FittsTaskOne (1D) and in Figure 17.6b for FittsTaskTwo (2D). (These are available as free downloads at http://www.yorku.ca/mack/FittsLawSoftware/. The downloads include Java source and class files, executable JAR files, examples, and detailed APIs.) For the 2D image, dashed lines are superimposed to show the sequence of target selections. As each target is selected, the highlight moves to a position across the layout circle to reveal the next target to the participant. Figure 17.6c shows a typical popup dialog after a sequence of trials using a mouse with FittsTaskTwo. The throughput value of 4.9 bits/s is typical for a mouse in this context.

ISO 9241-9 and the Fitts' paradigm have been used in many studies over the past 15 or so years. Examples of novel interactions or devices evaluated according to the standard include a trackball game controller (Natapov, Castellucci, & MacKenzie, 2009), smartphone touch input (MacKenzie, 2015), tabletop touch input (Sasangohar, MacKenzie, & Scott, 2009), *Wiimote* gun attachments (McArthur,

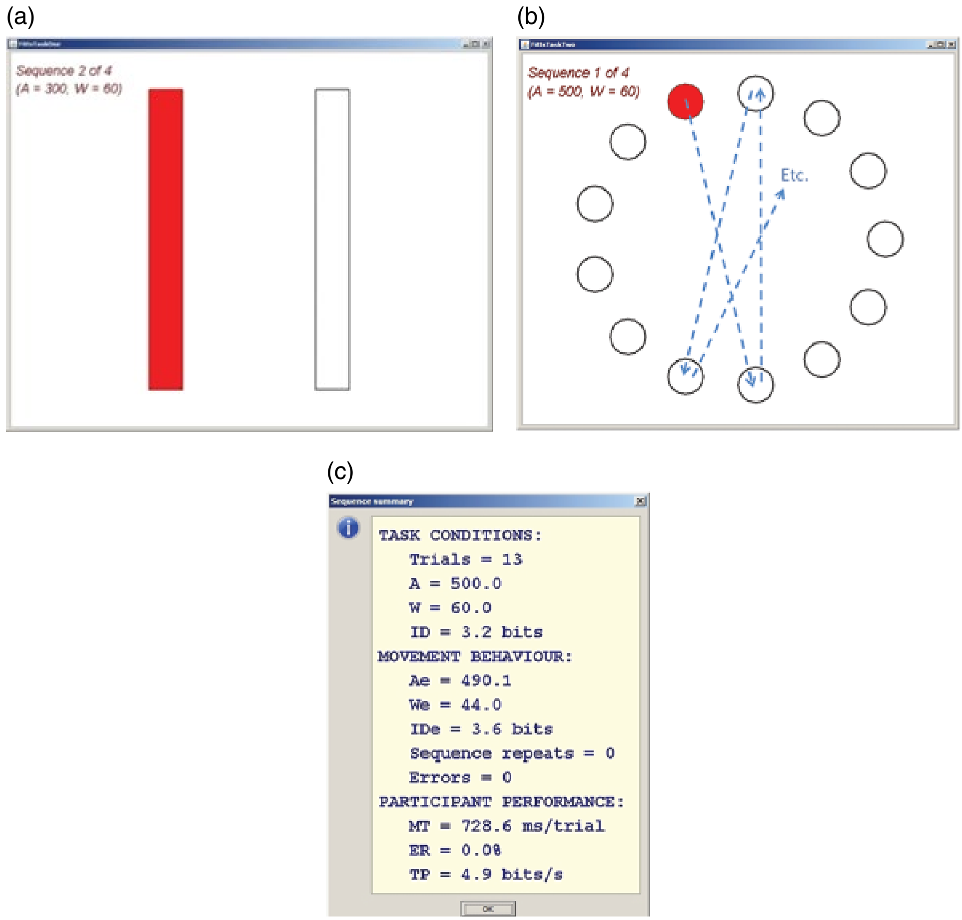


Figure 17.6 Implementations of the (a) one-dimensional (FittsTaskOne) and (b) two-dimensional (FittsTaskTwo) tasks in ISO 9241-9. (c) Popup dialog after a sequence of trials.

Castellucci, & MacKenzie, 2009), eye tracking (Zhang & MacKenzie, 2007), glove input (Calvo, Burnett, Finomore, & Perugini, 2012), and lip input (José & de Deus Lopes, 2015). Throughput values range from about 1 bit/s for lip input to about 7 bits/s for touch input. Mouse values are typically in the 4–5 bits/s range.

Calculation of Throughput

Although ISO 9241-9 provides the correct formula for Fitts' throughput, little guidance is offered on the data collection, data aggregation, or in performing the adjustment for accuracy. The latter presents a particular challenge when using the 2D task. In this section we examine the best practice method for calculating Fitts' throughput. We begin with Figure 17.7 which shows the formula for throughput, expanded to reveal the Shannon formulation for *ID* and the use of effective values for target amplitude and target width. The figure also highlights the presence of speed ($1/MT$) and accuracy (SD_x) in the calculation.

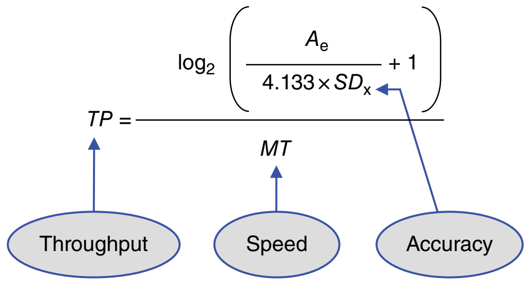


Figure 17.7 Formula for throughput showing the Shannon formulation for *ID* and the adjustment for accuracy. Speed (1/*MT*) and accuracy (SD_x) are featured.

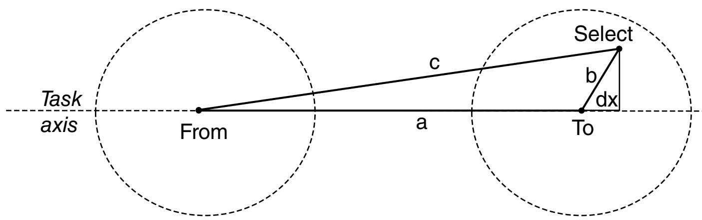


Figure 17.8 Geometry for a trial.

Whether using the 1D or the 2D task, the calculation of throughput requires Cartesian coordinate data for each trial. Data are required for three points: the starting position ("from"), the target position ("to"), and the trial-end position ("select"). See Figure 17.8. Although the figure shows a trial with horizontal movement to the right, the calculations described next are valid for movements in any direction or angle. Circular targets are shown to provide a conceptual visualization of the task. Other target shapes are possible, depending on the setup in the experiment.

The calculation begins by computing the length of the sides connecting the `from`, `to`, and `select` points in the figure. Using Java syntax:

```
double a = Math.hypot(x1−x2, y1−y2);
double b = Math.hypot(x−x2, y−y2);
double c = Math.hypot(x1−x, y1−y);
```

The *x*-*y* coordinates correspond to the `from` (x_1, y_1), `to` (x_2, y_2), and `select` (x, y) points in the figure. Given `a`, `b`, and `c`, as above, `dx` and `ae` are then calculated:

```
double dx = (c * c − b * b − a * a)/(2.0 * a);
double ae = a + dx;
```

Note that `dx` is 0 for a selection at the center of the target (as projected on the task axis), positive for a selection on the far side of center, and negative for a selection on the near side. It is an expected behavior that some selections will miss the target.

The effective target amplitude (A_e) is `ae` in the code above. It is the actual point-to-point movement distance for the trial, as projected on the task axis. For serial responses, an additional adjustment for A_e is to add `dx` from the previous trial (for all trials after the first). This is necessary since each trial begins at the selection point of the previous trial. For discrete responses, each trial begins at the center of the `from` target.

Given arrays for the `from`, `to`, and `select` points in a sequence of trials and the computed `ae` and `dx` for each trial, A_e is the mean of the `ae` values and SD_x is the standard deviation in the `dx` values. With these, ID_e is computed using Eq. 17.6 (substituting A_e and $W_e = 4.133 \times SD_x$) and throughput (TP) is computed using Eq. 17.3 (using ID_e). See also the equation in Figure 17.7. Of course, movement time (MT) is the mean of the times recorded for all trials in the sequence.

One final point concerns the *unit of analysis* for calculating throughput. The correct unit of analysis for throughput is an uninterrupted sequence of trials for a single participant. The premise for this is twofold:

- throughput cannot be calculated on a single trial;
- a sequence of trials is the smallest unit of action for which throughput can be attributed as a measure of performance.

On the first point, the calculation of throughput includes the variability in selection coordinates, akin to "noise." Thus, multiple selections are required and from the collected data the variability in the coordinates is computed.

The second point is of ecological concern. After a sequence of trials, the participant pauses, stretches, adjusts the apparatus, has a sip of tea, adjusts her position on a chair, or something. There is a demarcation between sequences and for no particular purpose other than to provide a break or pause, or perhaps to change to a different test condition. It is reasonable to assert that once a sequence is over, it is over! Behaviors were exhibited, observed, and measured and the next sequence is treated as a separate unit of action with separate performance measurements.

Given the above points, a closer look at the calculation of throughput is warranted. Consider Table 17.1. Each row in the table summarizes the results for 16 participants performing two 15-second sequences of trials at the indicated A and W. For each sequence, $MT = 15/m$, where m is the number of stylus taps. MT in the table is the mean computed over 16 participants, two sequences each. ID in the table is calculated from A and W using Eq. 17.2. Throughput for each row is calculated once, as ID/MT from the values in that row. The expanded formula for TP is as follows:

$$TP = \frac{\frac{1}{n}\sum_{i=1}^{n} ID_i}{\frac{1}{n}\sum_{i=1}^{n} MT_i} \tag{17.7}$$

where n is the number of participant × sequence combinations—32 in this case. But, the correct calculation, respecting the appropriate unit of analysis, is

$$TP = \frac{1}{n}\sum_{i=1}^{n} \frac{ID_i}{MT_i} \tag{17.8}$$

With Eq. 17.8, throughput is calculated on each sequence of trials. The overall throughput is the mean of n values. Equation 17.7 and Eq. 17.8 will yield the same value for the data in Table 17.1, because the iterated values for *ID* are the same across participants and sequences. But, when Crossman's adjustment for accuracy is used, the situation is different. The numerator in Eq. 17.7 is ID_e computed using W_e, as described earlier. Spatial variability is distilled into a single value, which in turn spawns a single ID_e. Let's call this ID_e'. Equation 17.7, with the adjustment for accuracy, is then

$$TP = \frac{ID_e'}{\frac{1}{n}\sum_{i=1}^{n} MT_i} \tag{17.9}$$

In essence, the accuracy component in a sequence of trials is differed. Accuracy is included at the end as a single composite adjustment applicable to all participants and trial sequences. Given the complexity of the log term for *ID*, this method is likely to introduce a bias in the calculation of throughput. Again, respecting the unit of analysis, the correct calculation for throughput including the adjustment for accuracy is

$$TP = \frac{1}{n}\sum_{i=1}^{n} \frac{IDe_i}{MT_i} \tag{17.10}$$

Equation 17.10 treats each sequence of trials as a separate unit of action. Speed and accuracy come together into a single measure of participant behavior, throughput. These measures are then summed and averaged across participants and trial sequences.

Equation 17.9 and Eq. 17.10 will yield different values for throughput. For the data in Table 17.1, *TP* = 8.97 bits/s using Eq. 17.9. This is in contrast to the value of *TP* = 10.10 bits/s seen in Table 17.1, which uses Eq. 17.7. It is not possible to recalculate throughput using Eq. 17.10 because the data from Fitts' experiment are not available for each participant on each trial sequence.

In summary, reducing the data from a Fitts' law experiment as in Table 17.1, while useful for summarizing participant responses or building a Fitts' law prediction equation (see Eq. 17.4), is not recommended if the goal is to measure the rate of information transfer (i.e., throughput; see Eq. 17.3 and Figure 17.7). For this, Eq. 17.10 should be used with each value for ID_e computed using Eq. 17.3 (as per Figure 17.7) on the data from a single sequence of trials. Here again we see a distinction between Fitts' law as a model for predicting, and Fitts' law as a model for measuring. Let's continue with an example of the use of Fitts' throughput for interactions typically found in contemporary computing systems.

Example User Study

We now put together the ideas above in an example user study investigating touch-based target selection on a smart phone.[2] Since 1D and 2D task types are both common in Fitts' law studies, it is worth asking whether there is an inherent difference in

[2] The example is a subset of a larger user study (see MacKenzie, 2015). The full study included an additional independent variable (device position: supported versus mobile) and additional dependent variables (movement time, error rate). The original study also examined results by participant finger size and tested the distribution characteristics of selection coordinates. Consult for details.

throughput for a 1D task compared to a 2D task. It seems this question has not been explored in a systematic way, that is, using task type (1D vs. 2D) as an independent variable in a controlled experiment.

Participants

Participants were recruited from the local university campus. The only stipulation was that participants were regular users of a touchscreen phone, pad, or tablet. Sixteen participants were recruited from a wide range for disciplines. Six were female. The mean age was 24.3 years ($SD = 3.0$). Participants' average touchscreen experience was 22.9 months ($SD = 15.8$). All participants were right handed.

Apparatus (hardware and software)

The testing device was an LG Nexus 4 touchscreen smartphone running Android OS version 4.2.2. The display was 61 × 102 mm (2.4 in × 4.0 in) with a resolution of 768 × 1184 pixels and a pixel density of 320 dpi. All communication with the phone was disabled during testing.

Custom Android software called FittsTouch was developed using Java SDK 1.6. The software implemented the serial 1D and 2D tasks commonly used in Fitts' law experiments and prescribed in ISO 9241-9. (FittsTouch is available as a free download including source code. See above.)

The same target amplitude and width conditions were used for both task types. The range was limited due to the small display and finger input. In all, six combinations were used: $A = \{ 156, 312, 624 \}$ pixels × $W = \{ 78, 130 \}$ pixels. These corresponded to task difficulties from $ID = 1.14$ bits to $ID = 3.17$ bits (see Eq. 17.6). A wider range is desirable but pilot testing revealed very high error rates for smaller targets. (This due to a phenomenon of touch input known as the *fat-finger problem*—Wigdor, Forlines, Baudisch, Barnwell, & Shen, 2007.) The scale of target conditions was chosen such that the widest condition (largest A, largest W) spanned the width of the display (portrait orientation) minus 10 pixels on each side. Examples of target conditions are shown in Figure 17.9.

The 2D conditions included 20 targets, which was the number of trials in a sequence. The target to select was highlighted. Upon selection, the highlight moved to the opposite target. Selections proceeded in a rotating pattern around the layout circle until all targets were selected. For the 1D task, selections were back and forth. Data collection for a sequence began on the first tap and ended after 20 target selections (21 taps).

Procedure

After signing a consent form, participants were briefed on the goals of the experiment. The experiment task was demonstrated to participants, after which they did a few practice sequences. They sat at a desk with the device positioned on the desk surface. They were allowed to anchor the device with their nondominant hand, if desired. An example of a participant performing trials in the 1D condition is shown

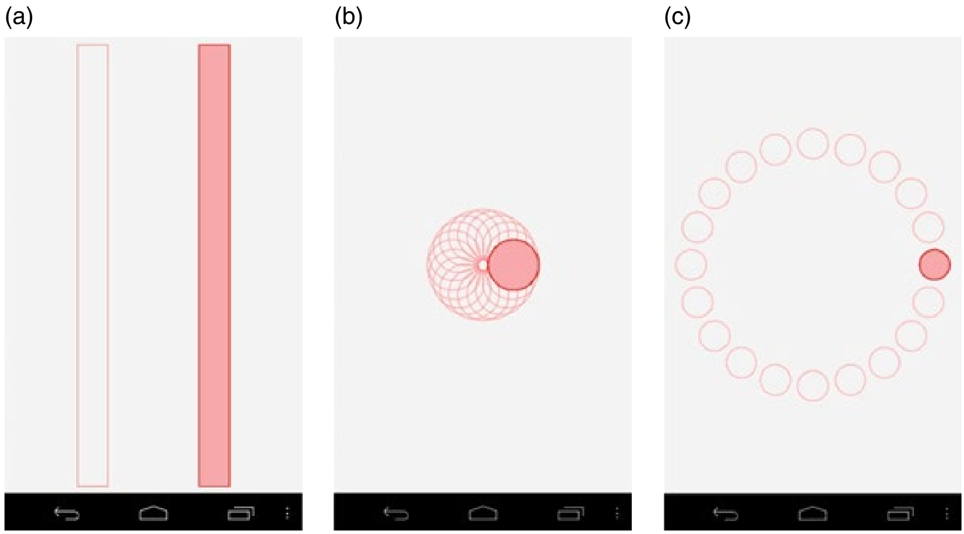


Figure 17.9 Example task conditions. (a) 1D with $A = 312$ & $W = 78$. (b) 2D with $A = 156$ & $W = 130$. (c) 2D with $A = 624$ & $W = 78$. All units pixels.

Figure 17.10 (a) A participant performing trials in the 1D condition. (b) Example dialog at the end of a sequence.

in Figure 17.10a. An auditory beep sounded if a target was missed. At the end of each sequence a dialog appeared showing summary results for the sequence. See Figure 17.10b for an example. The dialog is useful for demos and to help inform and motivate participants during testing.

Participants were asked to select targets as quickly and accurately as possible, at a comfortable pace. They were told that missing an occasional target was OK, but that if many targets were missed, they should slow down.

Design

The experiment was fully within subjects with the following independent variables and levels:

Task	1D, 2D
Block	1, 2, 3, 4, 5
Amplitude	156, 312, 624 pixels
Width	78, 130 pixels

The primary independent variable was task. Block, amplitude, and width were included to gather a sufficient quantity of data over a reasonable range of task difficulties.

For each condition, participants performed a sequence of 20 trials. The task conditions were counterbalanced with eight participants per order. The amplitude and width conditions were randomized within blocks.

The dependent variable was throughput. Testing lasted about 45 minutes per participant. The total number of trials was 16 participants × 2 tasks × 5 blocks × 3 amplitudes × 2 widths × 20 trials = 19,200.

Results and Discussion

The grand mean for throughput was 6.85 bits/s. This result, in itself, is remarkable. Here we see empirical evidence underpinning the tremendous success of contemporary touch-based interaction. Not only is the touch *experience* appealing, touch *performance* is measurably superior compared to traditional interaction techniques. For desktop interaction the mouse is well-known to perform best for most point-select interaction tasks. (A possible exception is the stylus. Performance with a stylus is generally as good as, or sometimes slightly better than, a mouse—see MacKenzie et al., 1991). In a review of Fitts' law studies following the ISO 9241-9 standard, throughput values for the mouse ranged from 3.7 bits/s to 4.9 bits/s (Soukoreff & MacKenzie, 2004, Table 5). The value just reported for touch input reveals a performance advantage for touch in the range of 40% to 85% compared to a mouse. (Of course, a direct comparison is not possible because mouse input is not supported on small touchscreen devices such the LG Nexus 4 used in this study.) The most likely reason lies in the distinguishing properties of *direct input* versus *indirect input*. With a mouse or other traditional pointing device, the user manipulates a device to indirectly control an on-screen tracking symbol. Selection requires pressing a button on the device. With touch input there is neither a tracking symbol nor a button: Input is direct!

The results for throughput by participant and task are shown in Table 17.2. The 1D task yielded a throughput of 7.43 bits/s, which was 18.5% higher than the mean of 6.27 bits/s for the 2D task. The difference was statistically significant ($F_{1,15} = 29.8$, $p < .0001$). All participants had higher throughput for the 1D task. Throughput was fairly flat over the five blocks of testing with < 3% change in throughput from block 1 to block 5. Consequently, a breakdown of results by block is not given.

The higher throughput for the 1D condition is explained as follows. With side-to-side movement only, the 1D condition is easier. Movements in the 2D condition are more complicated, since the direction of movement changes by 360°/20 = 18° with

Table 17.2 Throughput (bits/s) by participant and task.

	Task	
Participant	*1D*	*2D*
P01	6.28	6.19
P02	4.83	4.79
P03	5.90	5.34
P04	7.05	5.42
P05	7.83	5.83
P06	6.72	5.65
P07	6.38	5.05
P08	7.45	6.62
P09	8.26	6.09
P10	6.42	6.40
P11	8.33	5.94
P12	9.37	8.30
P13	8.75	6.17
P14	7.26	5.88
P15	9.01	7.76
P16	8.97	8.84
Mean	7.43	6.27
SD	1.30	1.13

each trial. Furthermore, occlusion is unavoidable for some trials in a sequence. This does not occur for the 1D task.

Throughput was calculated using Eq. 17.3 using the Shannon formulation for *ID* along with A_e and W_e (as per Figure 17.7). The unit of analysis for the calculation was a sequence of trials, as discussed earlier. Each value of throughput in Table 17.2 is therefore the mean of 30 values of throughput, since each participant performed five sequences of trials (1 per block) for each of six *A*-*W* conditions.

Figure 17.11 shows a chart of the findings for throughput by task, as might appear in a research paper. The error bars show ±1 *SD* using the values along the bottom row in Table 17.2.

Conclusion

This chapter has provided an overview of Fitts' law in view of current practice in human-computer interaction (HCI). It is important to bear in mind the long history of Fitts' law research in other fields and in the early years of HCI. In the 1950s, when Fitts proposed his model of human movement, graphical user interfaces and computer pointing devices did not exist. Yet, throughout the history of HCI (since Card et al., 1978), research on point-select computing tasks is inseparable from Fitts' law. The initial studies focused on device comparisons and model conformity. Since then—and partly due to the publication of ISO 9241-9—focus has shifted to the use of Fitts' throughput (in "bits/s") as a dependent variable. This is in keeping with Fitts' original intention to explore the information capacity of the human motor system.

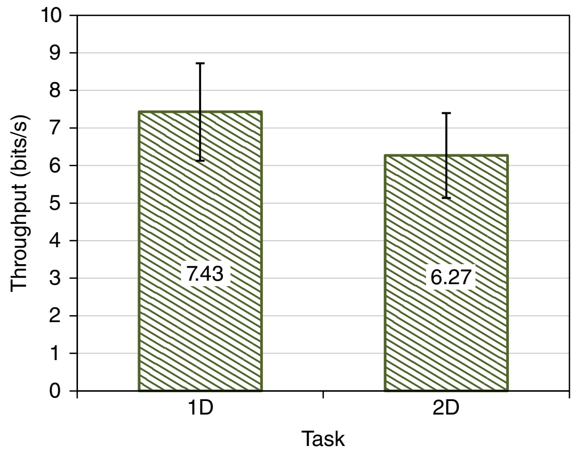


Figure 17.11 Throughput (bits/s) by task. Error bars show ±1 *SD*.

Much of this research has seen Fitts' law applied to topics only peripherally related to pointing devices. Examples include expanding targets, hidden targets, fish-eye targets, pointing on the move, eye tracking, force feedback, tilt input, gravity wells, multi-monitor displays, wearable computing, accessible computer, virtual reality, 3D, magic lenses, and so on. Research in these topics, and others, has thrived on the theory and information metaphor inspired and guided by Fitts' law. This is Fitts' legacy to research in human-computer interaction.

References

Calvo, A., Burnett, G., Finomore, V., & Perugini, S. (2012). The design, implementation, and evaluation of a pointing device for a wearable computer. *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting—HFES 2012* (pp. 521–525). Santa Monica, CA: HFES.

Card, S. K., English, W. K., & Burr, B. J. (1978). Evaluation of mouse, rate-controlled isometric joystick, step keys, and text keys for text selection on a CRT. *Ergonomics, 21*, 601–613.

Constantin, C., & MacKenzie, I. S. (2014). Tilt-controlled mobile games: Velocity-control versus position-control. *Proceedings of the 6th IEEE Consumer Electronics Society Games, Entertainment, Media Conference—IEEE-GEM 2014* (pp. 24–30). New York, NY: ACM.

Crossman, E. R. F. W., & Goodeve, P. J. (1983). Feedback control of hand-movement and Fitts' law: Communication to the experimental society. *Journal of Experimental Psychology, 35A*, 251–278.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, *47*, 381–391.

Fitts, P. M., & Peterson, J. R. (1964). Information capacity of discrete motor responses. *Journal of Experimental Psychology, 67*, 103–112.

Fitts, P. M., & Radford, B. K. (1966). Information capacity of discrete motor responses under different cognitive sets. *Journal of Experimental Psychology, 71*, 475–482.

Gillan, D. J., Holden, K., Adam, S., Rudisill, M., & Magee, L. (1990). How does Fitts' law fit pointing and dragging? *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems—CHI '90* (pp. 227–234). New York, NY: ACM.

Goldman, S. (1953). *Information theory.* New York, NY. Prentice-Hall.
Hick, W. E. (1952). On the rate of gain of information. *Quarterly Journal of Experimental Psychology, 4*, 11–36.
Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology, 45*, 188–196.
International Organization for Standardization. (2000). *Ergonomic requirements for office work with visual display terminals (VDTs)—Part 9: Requirements for non-keyboard input devices* (ISO 9241-9). Geneva: International Organization for Standardisation.
International Organization for Standardization. (2012). *Evaluation methods for the design of physical input devices* (ISO/TC 9241-411: 2012(E)). Geneva: International Organisation for Standardisation.
José, M. A., & de Deus Lopes, R. (2015). Human-computer interface controlled by the lip. *IEEE Journal of Biomedical and Health Informatics, 19*(1), 302–308.
MacKenzie, I. S. (1989). A note on the information-theoretic basis for Fitts' law. *Journal of Motor Behavior, 21*, 323–330.
MacKenzie, I. S. (1991). *Fitts' law as a performance model in human-computer interaction.* (Doctoral thesis, University of Toronto, Toronto). Retrieved from http://www.yorku.ca/mack/phd.html
MacKenzie, I. S. (1992). Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction, 7*, 91–139.
MacKenzie, I. S. (2012). Evaluating eye tracking systems for computer input. In P. Majaranta, H. Aoki, M. Donegan, D. W. Hansen, J. P. Hansen, A. Hyrskykari & K.-J. Räihä (Eds.), *Gaze interaction and applications of eye tracking: Advances in assistive technologies* (pp. 205–225). Hershey, PA: IGI Global.
MacKenzie, I. S. (2013). A note on the validity of the Shannon formulation for Fitts' index of difficulty. *Open Journal of Applied Science, 3*(6), 360–368.
MacKenzie, I. S. (2015). Fitts' throughput and the remarkable case of touch-based target selection. In *Proceedings of HCI International—HCII 2015 (LNCS 9170)* (pp. 238–249). Heidelberg: Springer.
MacKenzie, I. S., Sellen, A., & Buxton, W. (1991). A comparison of input devices in elemental pointing and dragging tasks. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems—CHI '91* (pp. 161–166). New York, NY: ACM.
MacKenzie, I. S., & Soukoreff, R. W. (2003). Card, English, and Burr (1978)—25 years later. *Extended Abstracts of the ACM SIGCHI Conference on Human Factors in Computing Systems—CHI 2003* (pp. 760–761). New York: ACM.
MacKenzie, I. S., & Teather, R. J. (2012). FittsTilt: The application of Fitts' law to tilt-based interaction. In *Proceedings of the 7th Nordic Conference on Human-Computer Interaction—NordiCHI 2012* (pp. 568–577). New York, NY: ACM.
McArthur, V., Castellucci, S. J., & MacKenzie, I. S. (2009). An empirical comparison of "Wiimode" gun attachments for pointing tasks. *Proceedings of the ACM Symposium on Engineering Interactive Computing Systems—EICS 2009* (pp. 203–208). New York, NY: ACM.
Meyer, D. E., Abrams, R. A., Kornblum, S., Wright, C. E., & Smith, J. E. K. (1988). Optimality in human motor performance: Ideal control of rapid aimed movements. *Psychological Review, 95*, 340–370.
Natapov, D., Castellucci, S. J., & MacKenzie, I. S. (2009). ISO 9241-9 evaluation of video game controllers. In *Proceedings of Graphics Interface 2009* (pp. 223–230). Toronto: CIPS.
Reza, F. M. (1961). *An introduction to information theory.* New York, NY: McGraw-Hill.
Sasangohar, F., MacKenzie, I. S., & Scott, S. (2009). Evaluation of mouse and touch input for a tabletop display using Fitts' reciprocal tapping task. In *Proceedings of the 53rd Annual Meeting of the Human Factors and Ergonomics Society—HFES 2009* (pp. 839–843). Santa Monica, CA: HFES.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communications.* Urbana, IL: University of Illinois Press.

Soukoreff, R. W., & MacKenzie, I. S. (2004). Towards a standard for pointing device evaluation: Perspectives on 27 years of Fitts' law research in HCI. *International Journal of Human-Computer Studies, 61*, 751–789.

Ware, C., & Mikaelian, H. H. (1987). An evaluation of an eye tracker as a device for computer input. In *Proceedings of the CHI+GI '87 Conference on Human Factors in Computing Systems and Graphics Interface* (pp. 183–188). New York, NY: ACM.

Welford, A. T. (1960). The measurement of sensory-motor performance: Survey and reappraisal of twelve years progress. *Ergonomics, 3*, 189–230.

Welford, A. T. (1968). *Fundamentals of skill.* London. Methuen.

Wigdor, D., Forlines, C., Baudisch, P., Barnwell, J., & Shen, C. (2007). Lucid touch: A see-through mobile device. In *Proceedings of the ACM Symposium on User Interface Software and Technology—UIST 2007* (pp. 269–278). New York, NY: ACM.

Zhang, X., & MacKenzie, I. S. (2007). Evaluating eye tracking with ISO 9241—Part 9. In *Proceedings of HCI International 2007* (pp. 779–788). Heidelberg: Springer.

18

Principles for Designing Body-Centered Auditory Feedback

Ana Tajadura-Jiménez, Aleksander Väljamäe, Frédéric Bevilacqua, and Nadia Bianchi-Berthouze

Scenario

Imagine yourself walking up the stairs. Suddenly, you hear someone rushing down the stairs from the upper floor towards you. Still not seeing the other person and just based on the footstep sounds, you can guess a number of things, such as the material of the staircase and the material of the stranger's shoe soles. You can somehow know whether the person has a small or a big body, perhaps also know whether the person is a he or a she, and you can even tell something about his or her emotional state. Now think about the sound of your own footsteps and imagine that it suddenly changes: instead of sounding as usual, they now rather sound as if produced by a much smaller body than yours. Would you feel small and light, like Alice in Wonderland after she drinks from the mysterious bottle? How do you think this would impact on your walking style or in your emotional state? And how do you think this event would impact on the perceived size of other objects and people around you or in the way you interact with them? Think about Alice and how the "little" door into the garden is not little anymore once Alice becomes small.

Introduction

For the majority of us, the feeling of having a body, moving it, experiencing its weight and the sense of touch when it interacts with other objects is so familiar that we do not usually think about it. Nevertheless, our body tightly connects us with the surrounding physical world. The example scenario above highlights a few of the phenomena related to this connection between our body and the physical (and possibly virtual) world. First, we interact with our surrounding environment through our body, and second, it is through our bodily senses, including hearing, that we perceive the world surrounding us (Damasio, 1999; Tsakiris & Haggard, 2005). Auditory feedback often accompanies our interactions with physical and virtual objects, interfaces or agents. The scenario above demonstrates how a person can perceive

The Wiley Handbook of Human Computer Interaction, Volume 1, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

various reality aspects just through footstep sounds. It can be information about the material of the ground surface (Giordano et al., 2012), the material of the shoes of the walkers (Giordano, Egerman, & Bresin, 2014), or about the movements and the body of the walkers (Li, Logan, & Pastore, 1991; Pastore, Flint, Gaston, & Solomon, 2008), including information about one's own body emotional and physical state (Tajadura-Jiménez et al., 2015a). The last point is especially important because our body posture often expresses our emotions, both intentionally and unconsciously (Kleinsmith & Bianchi-Berthouze, 2013). In addition, self-body perception highly impacts our self-esteem and emotional state (Carney, Cuddy, & Yap, 2010), forming a basis of self-identity (Giordano et al., 2014; Longo, Schuur, Kammers, Taskiris, & Haggard, 2008). Finally, the example scenario shows that one's body can be used as a "perceptual ruler" to measure the size of surrounding objects. Consequently, if one's body is perceived as smaller, by for example altering acoustic properties such as the reverberation time of the surrounding room, then external objects will be perceived as larger (like in *Alice in Wonderland*; Linkenauger, Witt, & Proffitt, 2011; van der Hoort, Guterstam, & Ehrsson, 2011).

Body-centered interaction

The central role of our body in perception, cognition and interaction, has been addressed by philosophers (e.g., Gallagher, 2005; Merleau-Ponty, 1962), psychologists (e.g., Niedenthal et al., 2005) and neuroscientists (e.g., Damasio, 1999; Damasio & Damasio, 2006; Tsakiris, 2010), and is often referred to as *embodied cognition*. This concept has also been applied in the context of human computer interaction (HCI) and in the design of multimodal interactive systems. More than two decades ago Slater and Usoh (1994) introduced a "body-centered interaction" paradigm that involves a number of components: (a) inference about the state of one's body; (b) body-centered feedback; and (c) magical and mundane interaction. Through a number of insightful and seminal experiments, Slater and colleagues showed that one's own body, either physical or virtual, plays a primary role when perceiving and interacting in immersive virtual environments (IVE) and it strongly influences the feeling of *presence* (i.e., "the sensation of being there," but also see more recent work on place and body illusions, e.g., Blanke, Slater, & Serino, 2015). In this perception-action loop (also known as sensory-motor loop or sensorimotor loop), data sensed by our body lead to the perception and cognition of the external reality, which in turn guide our next bodily actions, which influence the surrounding environment. This perception-action loop approach is reminiscent of the widely accepted neuroscientific theories of "forward internal models" of motor-to-sensory transformations (Wolpert & Ghahramani, 2000). These theories state that motor actions planning and execution are adjusted according to the match between the sensory feedback received when performing actions (*afferent* inputs) and the feedback predicted based on the signals generated by the motor system (*efferent copy*). In the theories of "forward internal models" the body takes a central place, because the predictions on the sensory feedback (including sounds) caused by our actions are based, among other factors, on the dimensions and configuration of our body through which we perform these actions. Similarly, in Slater and Usoh's view, presence in the actual or virtual reality occurs through our body becoming an object in that reality, and the perceptions of objects and agents are influenced by the way we mentally represent our bodies.

A mental model of our body guiding our interactions

Neuroscientific studies have repeatedly shown that our brain uses these body representations, or mental models of the body, when we move or touch objects. Body representations are indeed necessary for successful and smooth interactions with the environment (Head & Holmes, 1911–1912; Maravita & Iriki, 2004). They are continuously updated in response to sensory inputs received about the body (e.g., Botvinick & Cohen, 1998), but in a way can be considered *amodal* as they are not specific to a single sensory modality.

There are at least three different types of such mental body representations. First, there are representations of the physical appearance of the body, such as its shape and size, often known as "body image" (de Vignemont et al., 2010). Second, there are representations of the configuration and location of the different body parts in the space, often known as "body schema" (Haggard et al., 2006; Holmes & Spence, 2004). Third, there are representations of the region of space immediately surrounding the body, often known as "near space," "peripersonal space," or "personal space" (Lourenco, Longo, & Pathman, 2011; Tajadura-Jiménez et al., 2011).

It is often the case that mental body representations do not match exactly the actual physical appearance or configuration of the body. Most people experience some of these differences between actual and represented body, but sometimes the differences are extreme as in certain clinical disorders. Among such examples are anorexia nervosa, which is characterized by distortions in body image (Carruthers, 2008); some cases of chronic pain caused by the lack of awareness regarding one's body parts location in the space (i.e., proprioception; della Volpe et al., 2006); or distortions in the represented size of body parts, and peripersonal space (Lewis & McCabe, 2010). Such distortions impact on people's way of engaging in interactions with the environment, and with other people, and often also negatively impact on one's emotional state and self-esteem (Carruthers, 2008).

Altering representations of body and reality through sound interaction feedback

We described above how our perception of the environment and of our bodies is shaped by sensory feedback generated by our actions. For instance, we perceive the properties of the material of a surface, its roughness, hardness or coarseness, through the visual, haptic, and auditory cues received when touching it (Klatzky & Lederman, 2010). Similarly, we perceive our body through the continuous stream of multiple sensory inputs: proprioceptive, vestibular, tactile, visual, auditory, and so forth. So what happens if sounds generated by our body actions suddenly change? Will we perceive differently our body, the object involved in the interaction, or perhaps both? Evidence suggests that all of these can happen due to the fact that our representations of the environment and of our body are not fixed, and can be altered through sound interaction feedback.

Recent neuroscientific evidence suggests that object and body representations are continuously updated in response to auditory cues, similarly to the way in which they change through vision or touch. For instance, altering the frequency components of the sounds produced when touching a surface has been shown to change the perceived roughness of the touched surface (Guest et al., 2002). Similarly, altering the spectra and/or amplitude of the sounds produced when rubbing two hands together changes the perceived smoothness and dryness of the skin (Jousmäki & Hari, 1998).

The plasticity of these perceptions, together with the fact that current interactive multimodal technology allows these sounds to be altered and fed back to the listener, leads to many possibilities to change the perceptions people have of the world and of their own bodies. For instance, altering in real time the heard sound of people's own footsteps has been shown to change their perception of their own body size (Tajadura-Jiménez, 2015a). This possibility of changing the action sounds opens new avenues in the use and design of physical and virtual environments, objects, and agents; forms of interaction and body expressivity; and even ways of augmenting one's self-perceptions, with a great number of potential applications in all possible fields ranging from clinical neuroscience to entertainment.

Body perception, sound, and emotion

In this chapter we will often refer to emotional processes in relation to both body perception and sound. Indeed, sounds often elicit emotional responses in listeners as exemplified in our opening scenarios and as demonstrated by many studies on the impact of a broad range of sounds on emotional states (Bradley & Lang, 1999; Juslin & Västfjäll, 2008; Lenti Boero & Bottoni, 2008). Our sensory systems are responsible for keeping a constant margin of safety surrounding our body, and in this respect, the most basic function of the auditory system is to act as a warning system by eliciting emotional responses (e.g., Juslin & Västfjäll, 2008). Therefore, by hearing, we are able to monitor the surrounding environment in order to detect the presence of other individuals (or objects) and to obtain information about them (e.g., localization, size or material), their actions (e.g., direction or velocity), and, if needed, possible safe escape routes (Hermann & Ritter, 2004).

Emotional responses are triggered by the information gathered through the sound, and changes in one's body physiological state occur in order to get the body ready for action (e.g., Ekman, 1980; LeDoux, 1998; Levenson, 1994). While people such as James (1894) have argued that emotions arise from cognitive interpretations of these physiological changes in response to events, it is now accepted that emotions are psychological states characterized by behavioral, visceral, and experiential changes (Seth, 2013). They affect attention, cognitive, and perceptual processes (de Gelder & Vroomen, 2000) and influence our judgments and decisions (Peters et al., 2006).

Emotional processes will either pull our bodies away from the sound source, in case of danger, or push our bodies towards the sound sources in case of feeling attraction towards them. This highlights the central role of body perception in relation to emotional responses to events, since one's body is taken as a reference frame in establishing one's position in relation to the objects, individuals, events, and situations around us and reacting emotionally to them (e.g., Damasio, 1999; James, 1890). Given that one's body perception may modulate emotional responses to sound, and that at the same time sound can also induce emotion and changes in body perception, the interaction between these three aspects needs to be considered in a body-centered sound design.

Body-centered sound design: Outline of this chapter

Although many of the ideas discussed in this chapter apply to all other sensory feedback, we focus on sound cues (and related sound design solutions) that provide critical information about body-centered interaction; this is a subset of sensory cues rather neglected in HCI. We selected recent research from multiple areas including virtual reality (VR), sports, health, rehabilitation, and art from the perspective of body-centered sound design and how such design is and can be informed by recent neuroscience and psychology findings. Thus, our focus is not specifically on technology or on sound synthesis but rather on the top-down and bottom-up brain processes that can guide the effectiveness of sound-based body-centered interaction design.

In this chapter, we discuss how neuroscientifically grounded insights contribute to the design of new, enhanced interactive technologies. The chapter lists a number of interesting advantages of using sound for HCI applications and then it focuses on how auditory cues can be used to enhance the perception of self by constructing a sound-based representation of one's own body, a sonic self-avatar, in mixed reality environments. The various types of cues for representing sonic avatars, like footsteps, heartbeat or voice, and for representing virtual space around the body (one's "safety margin" or near space), are discussed. Next, we concentrate on examples of altering our body representation by sound to enhance positive emotions and facilitate movement and motor learning. We then present examples in which sound is used to alter the properties of objects being touched in order to also alter movement dynamics and emotional state. Finally, we discuss how new applications based on these principles can be used in rehabilitation and, in particular, to help to overcome physical and psychological barriers in people with specific physical or psychological conditions.

We deliberately decided to narrow the topic of this chapter to "designing through body-centered sound" due to its emergent nature and actuality of embodied cognition themes. However, there are many thorough works that concern topics closely related to this chapter. For general sonic interaction design we refer the reader to the book of the same title edited by Franinovic and Serafin (2013). Very inspiring treatments of the topic of music and emotion are offered in the article by Juslin and Västfjäll (2008) and the handbook by Sloboda (2010). The topic of music also closely relates to embodied music cognition, including action-based effects on music perception, and interpersonal synchrony and entrainment by listening to music or by playing music with other people (Delaherche et al., 2012; Leman, 2007; Maes, Leman, Palmer, & Wanderley, 2014; Reidsma, Radha, & Nijholt, 2014; Zatorre, Chen, & Penhune, 2007). For an overview of technologies related to virtual and mixed auditory realities and presence we suggest Larsson, Väljamäe, Västfjäll, Tajadura-Jiménez, & Kleiner (2010) and references therein. While we concentrated on sound in this chapter, our perception is multisensory and we invite interested readers to learn about multisensory design principles in the article by Soto-Faraco and Väljamäe (2012).

Self-representation via Sound in Mixed Reality Environments

According to the "body-centered interaction" paradigm introduced by Slater & Usoh (1994), having a representation of virtual body states, as well as having body-centered sensory feedback (e.g., visible parts of a virtual body), are crucial components for high presence responses and engagement in virtual environment (VE) and mixed reality applications. In this section we address some of the examples that concern such a virtual sonic body and acoustical surroundings.

Advantages of using sound for HCI applications

The use of sound offers a number of interesting advantages for HCI applications. First, it does not interfere with movement. Second, it can potentially inform the user of events outside of their visual field since audition has a 360 degrees field of view. It has also been shown that one can create a space representation based on sensorimotor and auditory cues only (Viaud-Delmon & Warusfel, 2014). Third, the use of sound allows for the presentation of several streams of information (e.g., tactile and gesture) in parallel (Hermann & Ritter, 2004). And finally, it can provide a continuous flow of information, as audition never "turns off" in the same way that vision is blocked when shutting our eyes (Larsson, 2005), not mentioning the assistive technologies for visually impaired (Csapó, Wersényi, Nagy, & Stockman, 2015). A further advantage of using sound is that the auditory system operates relatively well, even in noisy environments, offering a high temporal resolution and a high sensitivity for detecting structured motion (e.g., rhythm; Hermann & Ritter, 2004). As a matter of fact, audition has been characterized as a change detector that responds to certain sound properties indicating a rapid change. It does so by quickly orienting behavior towards the changes that may signal potential threats (c.f. Juslin & Västfjäll, 2008); this response is faster than that observed for the visual system (McDonald, Teder-Salejarvi, & Hillyard, 2000). Hence, incorporating sounds that trigger intuitive, fast and accurate responses in users, might be beneficial for the design of systems where sound is used to convey alerts and warnings, such as vehicles, emergency systems in hospitals, or working environments (for recent reviews see Edworthy, 2013; Roginska, 2013). Finally, from the point of view of aesthetics, sound is also a source of enjoyment and entertainment (Juslin & Västfjäll, 2008).

Sonic self-avatar

Our body produces many different sounds due to its own internal activity and to its interaction with the environment. Think, for instance, of the sounds produced by one breathing, yawning, chewing, walking or tapping on surfaces. These sounds are rich in information about one's body, its dimensions, its emotional state, the location of the body parts, and the progress of the actions being executed. People do not pay attention to these sounds all the time, but they accompany us constantly, and can be considered a continuous "soundtrack" in our lives. People report that when these sounds are missing, for instance, when wearing earplugs, they have an altered awareness of their own bodies (Murray, Arnold, & Thornton, 2000). Further, when hearing breathing or heartbeat sounds, people unconsciously tend to associate these sounds with their own physiological state, and these sounds have effects on the cognitive and emotional processes of the listener (Phillips, Jones, Rieger, & Snell, 1999; Tajadura-Jiménez, Väljamäe, & Västfjäll, 2008). For instance, adding naturally breath intake sounds to synthetic speech seems to aid listeners to recall sentences

(Whalen, Hoequist, & Sheffert, 1995). Similarly, listening to heartbeat sounds with fast or slow rate can respectively increase or decrease emotional arousal (e.g., Woll & Mcfall, 1979), as well as impact on participants' own heartbeat rate and recall of emotional events (Tajadura-Jiménez et al., 2008). Moreover, eating has been described as an emotional experience that involves "being aware of and listening to the crunch of each bite and noise of the chewing sound in your head" (Albers, 2003).

Without the sounds our body produces when interacting with the environment it would be difficult to coordinate apparently simple actions such as brushing our teeth, moving objects from one position to another, turning the car on, or plugging in an electrical appliance. For instance, the motor behavior when touching a surface with the hand continuously "sculpts" the feedback sound, and vice versa. These sounds also provide feedback about the objects we are interacting with and about one's own body. For instance, the sounds produced when walking on a ground surface depend on the footgear and ground material, but also on the walker's weight and walking rate (Visell et al., 2009). Hence, these sounds can be used to enhance the perception of self in IVEs. Virtual reality research suggests that presence in IVEs occurs through our body becoming an object in that virtual reality. It has been reported that seeing a self-avatar—that is seeing, for example, our virtual body from a first-person perspective by looking down through a head-mounted display or by looking at a virtual mirror—increases presence in IVEs (Slater & Usoh, 1994; Dodds, Mohler, & Bülthoff, 2011). Similarly, one could think of having a sonic avatar body—this is an avatar that is not visual but it is constituted by bodily sounds and interaction sounds and which represents one's body in these IVEs. As with the visual avatar, we now know that adding sounds representing our body moving in the IVEs increases navigation and interaction in those IVEs (Nordahl, 2006). Interesting examples of sonic self-avatars can be found in some audio-only games (see for instance www.blindsidegame.com).

Following this "body-centered interaction" paradigm, the addition of sounds representing one's self-motion, such as an engine-like sound added to the moving auditory scene, has been shown to increase significantly the sensation of movement in IVEs in a study by Väljamäe and colleagues (Väljamäe, Larsson, Västfjäll & Kleiner, 2008a). In this study, the engine sound, unlike other sounds in the moving virtual scene, was used as a stationary anchor, staying on the first-person "point of hearing" (see Figure 18.1). This study also assessed mental motor imagery, which refers to users' abilities to imagine dynamic processes, including self-motion (Hall, Pongrac, & Buckholz, 1985). It is common to separate motor imagery into visual motor imagery (visualizing the body performing a movement) and kinesthetic motor imagery (imagining the feeling that the actual body movement produces). The engine sound facilitation effect was significantly correlated with participants' kinesthetic imagery but not with visual or auditory imagery, thus suggesting the relation of a first-person perspective in the perception of the self-motion representation sounds. For a review of other auditory cues influencing sensation of self-motion see Väljamäe (2009).

Further, when thinking about the constitution of a sonic self-avatar it is also important to consider the sound of one's avatar voice, as the acoustic cues of people's voices vary greatly across individuals and are therefore very important in representing one's self, for example, in game applications (Wadley, Carter, & Gibbs, 2014). There are voice acoustic cues that are usually good indicators of the age, gender, and size of the

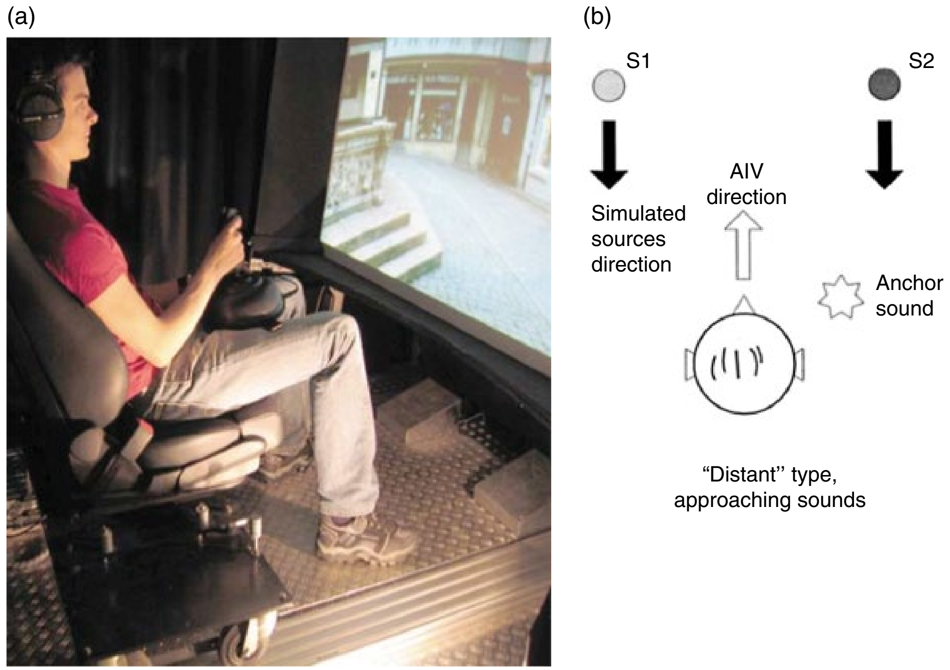


Figure 18.1 Having a sonic self-avatar in virtual immersive environments (left panel); schematic representation of auditory scene with the presence of the engine, anchor sound (right panel). AIV = Auditory-induced vection.

speakers (for a review on salient cues in voice signals see Smith & Patterson, 2005). Hence, when constituting a self-avatar of a particular age, gender or body size, one could think about voice morphing techniques that manipulate people's voices, as they talk, and present the resulting sounds from a first person perspective (Deutschmann, Steinvall, & Lagerström, 2011). Note that when people hear their own voice, approximately 50% of the sound energy they get is by air conduction through their ears and the other 50% of the sound energy is transmitted through bone conduction (Pörschmann, 2000; Väljamäe et al., 2008b). Hence, an effective presentation of the voice in this case may combine rendering of the voice through headphones (air conduction) and through a bone-conduction headset.

Finally, in the study on heartbeat sounds mentioned above, it was observed that the effects of heartbeat sounds on people's emotional state are heightened when the sounds are presented from close distance to the listener, either by using headphones instead of distant loudspeakers or by using vibrations to "capture" the sounds towards the listener (Caclin, Soto-Faraco, Kingstone, & Spence, 2002; Tajadura-Jiménez et al., 2008). This finding may suggest the importance of this first-person perspective of bodily sounds for the constitution of a sonic self-avatar (i.e., for sounds to represent one's body in the environment), paralleling the findings showing the importance of seeing a self-avatar from a first person perspective. These findings also highlight different effects between presenting sounds in the near space (or peripersonal space) and in the space far from the virtual or actual body of the listener. We expand on research on these differences in the next subsection.

Peripersonal space and acoustics in interactive virtual environments (IVEs)

We already mentioned that one of the types of mental body representations is representations of the region of space immediately surrounding the body, often known as "near space," "peripersonal space" (also referred to as PPS) or "personal space" (Lourenco et al., 2011; Tajadura-Jiménez, Pantelidou, Rebacz, Västfjäll, & Tsakiris, 2011). Neuroscientific and psychological studies have shown that our brains are specialized for the processing and integration of sensory information in the near space, as compared to the far space (e.g., Brozzoli, Gentile, Petkova, & Ehrsson, 2011; Rizzolatti, Scandolara, Matelli, & Gentilucci, 1981). It should be also noted that spatial resolution and sensitivity of our auditory processing is different for the so called far field (>1–1.5 m) and the near field, which is also reflected in different technological solutions to render acoustical events in a virtual space (e.g., Zotkin, Duraiswami, & Davis, 2004; see also Larsson et al., 2010, for more details and references).

In relation to external sound events and PPS, for instance, Neuhoff, Planisek, and Seifritz (2009) demonstrated that the terminal distance of sound sources approaching one's body is underestimated and Tajadura-Jiménez et al. (2009) showed that the integration of sound and touch is facilitated for sound events near (versus far from) the body. Several explanations have been offered for this specialization and for the importance of maintaining a representation of this space immediately surrounding the body, including the necessity of keeping track of objects located near the body in order to interact with them successfully (Chieffi, Fogassi, Gallese, & Gentilucci, 1992; Moseley, Gallace, & Spence, 2012) but also for maintaining a margin of safety around one's body (Graziano & Cooke, 2006; Niedenthal, 2007).

There is also a tight link between PPS and emotion. Indeed, in the field of social psychology the term "personal space" is often preferred over PPS and it is used to define the emotionally tinged zone around the human body that people experience as "their space" (Sommer, 1959), into which others cannot intrude without arousing discomfort (Hayduk, 1983). Research on auditory-induced emotion has found that the spatial positioning of sound sources in relation to the listener's PPS might modulate the emotional responses to the sound. In particular, more intense emotional responses are observed in relation to unpleasant sound sources perceived to be approaching (Ferri, Tajadura-Jiménez, Väljamäe, Vastano, & Costantini, 2015; Taffou & Viaud-Delmon, 2014; Tajadura-Jiménez, Väljamäe, Asutay & Västfjäll, 2010b) or close to the listener's body (Tajadura-Jiménez et al., 2008), or perceived as being behind the listener (Tajadura-Jiménez, Larsson, Väljamäe, Västfjäll, & Kleiner, 2010a; Asutay & Västfjäll, 2015) as compared to sources that are far, static, receding, or in the front space. Furthermore, the acoustically perceived dimensions of the surrounding space around one's body can impact on the emotional responses to sound events (e.g., being in a small virtual room with a lion), as in a way they set the *context* in which events occur and therefore influence our interpretation of them (Tajadura-Jiménez et al., 2010a).

The presented evidence can be summarized in design principles, including that sounds in the near space (a) are preferentially and differently processed than those sounds in the far space, (b) have greater chances to constitute a sonic self-avatar, and (c) can induce more intense emotional responses in listeners.

Altering Body Representation via Sound to Enhance Body Capabilities

In previous sections we already introduced the findings from neuroscience research that our brains use the available sensory inputs, including sound, to keep track of the continuously changing appearance of our body ("body image"), the configuration and position of our different body parts in space ("body schema"), and the space immediately surrounding our body ("peripersonal space"), as well as to keep track and adjust the motor actions (Botvinick & Cohen, 1998; De Vignemont, Ehrsson, & Haggard, 2005; de Vignemont, 2010; Haggard, Kitadono, Press, & Taylor-Clarke, 2006; Holmes & Spence, 2004; Lourenco, Longo & Pathman, 2011; Tajadura-Jiménez et al., 2011; Wolpert & Ghahramani, 2000). We also introduced the theories of "forward internal models" of motor-to-sensory transformations (Wolpert & Ghahramani, 2000), which suggest that we predict the auditory feedback we should receive from our bodily interactions (e.g., the sound produced when tapping a surface with one's hand) by considering the motor commands our brains sent (e.g., the applied force, direction of movement) and the mental model of our body dimensions and configuration. When the sound feedback received does not match these predictions, an update of the internal body models may occur, and this will also result in adjustments of the subsequent actions we perform. In this section we will briefly review the findings from neuroscience research on the effects of sound linked to body actions in the planning and adjustment of subsequent actions, as well as in the body image and body schema. We will then focus on the opportunities and challenges that these findings open for the design of technology in different contexts, including sports, health, and rehabilitation. Recent HCI research, building on neuroscience findings, is starting to exploit the use of sound linked to body actions for enhancing the perceived body and its capabilities.

Action-related sounds are used to plan and guide own actions

Neuroscience research shows a tight connection between perception of action-related sounds and the brain motor commands for real or imagined action. For instance, hearing sounds produced when performing certain actions, such as tearing a paper into pieces, activates the same brain areas that would have been recruited when preparing to perform the same actions (Aglioti & Pazzaglia, 2010; see related work on *mirror neurons*, which are involved in action execution and action observation, e.g., Rizzolatti & Craighero, 2004 or Keysers, 2010; see also related work on *motor contagion* mentioned in Chapter 28 of this handbook). Similarly, listening to sounds related to one's own actions can have an effect on the execution of subsequent actions. For instance, delaying the sounds produced by walking results in adjustments of one's gait (Menzer et al., 2010), and real-time alteration of the sound cues related to the strength applied when tapping a surface results in adjustments in the subsequently performed tapping movements (Tajadura-Jiménez et al., 2015b). The latter study also showed that emotional experience is affected by the congruence between tapping sounds and tapping actions. When hearing a *weak* tapping sound that did not match the sound predicted according to the applied strength of tapping, people felt less pleasant, more aroused, and less able to tap as compared with a condition where tapping actions and sounds were kept congruent. Interestingly, in the condition where people heard a *strong* tapping sound, the mismatch with the tapping action was less evident and pleasanter than in the *weak* tapping sound condition; in the *strong* condition participants felt able to tap and pleased with the sound feedback from their action, and they did not adjust their tapping behavior as much as in the *weak* condition. These findings demonstrate that people use body-related sounds to guide their own actions, which opens possibilities for technology to exploit these processes to facilitate and guide movement by using sound (Tajadura-Jiménez et al., 2015b). Further, the tight link between auditory and motor areas in the brain extends to other types of sounds, such as rhythmic sounds, which are often used to entrain movement (Kenyon & Thaut, 2005; Schneider, Münte, Rodriguez-Fornells, Sailer, & Altenmüller, 2010; see also work on interpersonal synchrony and entrainment by playing or listening to music, mentioned in the introduction: Delaherche et al., 2012; Reidsma et al., 2014).

Altering mental body models via sound

A few recent studies have shown that altering in real time the action sounds made by one's body can have an effect on the perceived body size. For instance, altering the spatial location of the sounds that are produced when one's own hand taps a surface, so that the heard sounds originate at double the distance at which one is actually tapping, can lead to the perception of one's arm being longer than before (Tajadura-Jiménez et al., 2012; Tajadura-Jiménez, Tsakiris, Marquardt, & Bianchi-Berthouze, 2015c), as well as to perform reaching actions as if one's arm was longer (Tajadura-Jiménez, Marquardt, Swapp, Kitagawa, & Bianchi-Berthouze, 2016). Importantly, in order to observe a recalibration of the body image and body schema triggered by these effects, the spatial manipulations need to be kept within certain spatial and temporal limits. The sound location should allow for being perceived as coming from one's body, and the presented sounds need to be in synchrony with one's actions. Our work shows that presenting the tapping sounds at quadruple the distance at which one is actually tapping, or in temporal asynchrony over 300 ms with one's action did not have effects on the perceived body size (Tajadura-Jiménez et al., 2012; Tajadura-Jiménez et al., 2015c; see also Lewkowicz, 1996, 1999 for research on the multisensory integration window during which asynchronous stimuli in different modalities are perceived as simultaneous). Another study showed that altering the frequency spectra of the sounds one's body produce when walking, so that the resulting sounds are consistent with those produced by a lighter body, can alter the perception of one's body size and weight, so that one's body feels thinner and lighter (Tajadura-Jiménez, 2015a; see also Tonetto, Klanovicz, & Spence, 2014). This study also showed that walking sounds making people feel thinner were connected to changes towards more active walking behaviors and more positive emotional states. These findings demonstrate that action-related sounds can have an effect on people's body image and body schema, which again opens possibilities for technology to exploit these processes to make people more aware of their body or to make them feel in a different body.

Sound for sensory substitution or sensory enhancement of body perception

A body-centered design exploiting these findings could use sound, first, as a source of information about the *actual* body dimensions, position, and movement—the use of sound in this context can be seen as a sort of *sensory substitution*. Several studies have validated the use of sound to give information that is normally channeled through touch.

For instance, in the study by Boyer et al. (2015), a sound is heard when "touching" a virtual surface in the air. The actual curvature of the surface can be estimated through the auditory feedback when exploring the surface. While the average precision remains lower than published results using touch, this study showed that the auditory feedback could be used effectively to provide participants with subtle spatial information of a virtual object. Second, sound could also be used to provide altered information about the body dimensions, position, and movements—the use of sound in this context can be seen as a sort of *sensory alteration* or *sensory enhancement*. There is a growing body of work supporting the use of sound in both contexts to increase positive body awareness and facilitate movement execution and engagement, as being more aware of one's body or feeling in a different body may affect emotional state and the planning of subsequent actions. For example, Boyer showed that sonification could enhance the accuracy in a visuo-manual tracking task (Boyer, 2015). As we already mentioned in the introduction, there is also a tight connection between sound and emotional state, which could be exploited in these applications (see above, and also Bradley & Lang, 1999; Juslin & Västfjäll, 2008; Lenti Boero & Bottoni, 2008; Tajadura-Jiménez et al., 2008, 2010a, 2010b), as well as the effects of auditory-induced emotion on motor behavior (Leman et al., 2013). A body-centered sound design exploiting these findings could thus have applications in VR, art, and entertainment contexts seeking to make people feel in a different body and to enhance a body-related emotional state, but also applications in health and rehabilitations contexts seeking to increase positivity about one's body and facilitate action.

The use of sound as a source of information about the *actual* body dimensions, position, and movement, has been shown to have positive effects in sports, dance, motor learning, and rehabilitation contexts, such as enhancing body awareness and movement coordination, increasing motivation, reducing anxiety related to physical performance, and enhancing the emotional state related to one's body (Großhauser, 2012; Rosati, Rodà, Avanzini, & Masiero, 2013; Schaffert, Mattes, & Effenberg, 2010; Singh et al., 2014; and Vogt, Pirro, & Kobenz, 2009). For instance, sound feedback can facilitate motor learning by guiding movement by providing information such as the distance to a target posture (Sigrist, Rauter, Riener, & Wolf, 2013), and can also improve self-efficacy (Singh et al., 2014). For a recent review on sensorimotor learning with movement sonification see the article by Bevilacqua et al. (2016). Much work has been done in the area of physical rehabilitation using sound as a source of physical information or to address psychological barriers. Much less work has been done using sound as a source of *sensory alteration* of one's own body perception. It was only recently that the possibility of altering one's own body perception with sound, to enhance physical performance, self-esteem, and positive attention to one's body has been explored in the context of HCI. This is the case of the study that altered walking sounds in real time to make people feel thinner and lighter, which showed that these changes in body perception made people walk "lighter" and induced a more positive emotional state (Tajadura-Jiménez et al., 2015a). Future research may consider applying similar body-centered sound design principles.

Finally, it is worth noting that the use of interactive sound feedback as a source of information about one's body has been experimented in performing arts, such as music and dance (Bevilacqua, Schnell, Fdili Alaoui, Klein, & Noeth, 2011). The case of dance is particularly interesting because dancers are particularly trained to perceive their own body and space. For example, Viaud-Delmon et al. (2011) reported a study where a dancer experimented with an interactive system where sound and music were responding to her position and movements. This study evaluated whether such a system could alter dancers' perceptions associated with movements and extend their body space. The dancer in this study reported that the system helped to increase her physical awareness, and induced a change of her perception of the stage. In another study, Francoise, Fdili Alaoui, Schiphorst, and Bevilacqua (2014) reported qualitative results, where dancers used an interactive system that was tuned to provide continuous auditory feedback to specific movement qualities. In this case, the sound was based on previously recorded vocalization of the dancers. They reported that the system tended to induce the exploration of different movement and modify their behaviors. In particular, the dancer felt that such a system could create a reflective space for movement learning. These studies clearly show that more research is needed in this promising area, to better establish how sound interactive systems could enhance body and movement perceptions.

Altering Interactions with Objects and Multimedia Interfaces via Sound to Enhance Movement Dynamics and Emotional State

As already said, sound is of primary importance to inform us about our surrounding environment, both because of the omnidirectional nature of hearing and because the main function of the auditory system is to act as a warning system. Several types of listening have been proposed (Caramiaux, Altavilla, Pobiner, & Tanaka, 2015), such as the *music listening* that focuses on elements such as pitches, timbre, and rhythm, or *everyday listening focusing on events*, as proposed by Gaver (1993). In the case of a sound produced by a specific action, it has been shown that listeners have the tendency to describe the sound by the action itself (Lemaître & Heller, 2013). Moreover, we also perceive the size and material of the objects though the sound (Giordano & McAdams, 2006; Grassi, Pastore, & Lemaître, 2013; Klatzky, Pai, & Krotkov, 2000).

A material perception—action loop during interactions with objects

The case of sound produced by objects through bodily interactions represents a specific case of body perception through sound. As our interactions with objects generally produce sounds, we constantly use the auditory feedback to adapt our movements as we touch an object to obtain information about its material. Interestingly, as in the case of our body perception, we are rarely conscious of how important these auditory feedback mechanisms are for our interaction through objects. For example, Cabe and Pittenger (2000) showed that users could successfully fill a vessel with water by using only auditory feedback. In this case the movement is continuously controlled through an action-perception loop (or vice versa, perception-action loop), where anticipation and adaptation come into play.

The association between our actions on objects and the resulting sounds is learned through our experience (Lemaître, Heller, Navolio, & Zúñiga-Peñaranda, 2015) and new associations can be learned. For example, the study by Cabe and Pittenger (2000) could lead to different results if different liquids causing different "filling" sounds, unknown to the participants, would be used. It has been also shown that the sound can activate motor representation of the movement. Several authors investigated movement that can be associated to sound or "sonic gestures." For example, Caramiaux, Schnell, Françoise, and Bevilacqua (2014) studied different movement strategies that users might use when asked to describe gesturally a sound while they are listening to it. In particular, they showed that identification of the action causing the sound favors the mimicking of the action.

The sound produced when touching an object can also be used to plan and execute subsequent movements. For instance, Castiello and colleagues showed that the speed of reaching-to-grasp movements (reaching towards an object) can be modulated by hearing, at movement onset, the sound that will be produced when grasping the object (as opposed to hearing the sound produced when grasping an object of a different material; Castiello, Giordano, Begliomini, Anuini, & Grassi, 2010). Several research studies also showed that the perceived floor material though sound feedback can influence the walking style (Bresin, de Witt, Papetti, Civolani, & Fontana, 2010; Tajadura-Jiménez et al., 2015a; Turchet & Bresin, 2015).

Changing sound feedback on object properties to alter movement dynamics

Similarly to the notion of affordance initially proposed by Gibson (1986) in the ecological approach of visual perception, several authors extended the notions of affordance to the interaction between sound and actions (Gaver, 1991; Godøy, 2009; Altavilla, Caramiaux, & Tanaka, 2013). Such notions of affordances have been used as design principles for the building of various interactive objects (Caramiaux et al., 2015; Houix, Misdariis, Susini, Bevilacqua, & Gutierrez, 2014), where the action and sound are coupled. Importantly, the use of multimodal interactive systems allows for designing the relationships between the actions and the sound feedback. The fact that we can easily, in a quantifiable manner, alter the audio feedback in response to the object manipulation opens opportunities to study our perception of the interaction (see Figure 18.2).

For example, Tajadura-Jiménez and colleagues described an interactive system that allows altering the perception-action loop between the touch of a surface, and the sound produced through the action (Tajadura-Jiménez, Liu, Berthouze, & Bevilacqua, 2014). In this study, the original sound of the touch was modified through digital sound processing, creating in turn the illusion of different surface textures such as grainy or smooth surfaces. We showed that by altering the sound feedback, both the touching finger movement dynamics and the material perception could be modified. This result is in line with the hypothesis that such an interactive system can modify the perception-action loop at the perception and motor levels.

Emotion is also intricated in this material perception–action loop. Changing the auditory information about motor performance or about the object of interaction may impact on how we feel about our motor capabilities or about being interacting with an object of particular characteristics. Using an interactive system, Tajadura-Jiménez and colleagues studied the influence of sound feedback to tapping either on a physical object (e.g., a table) or a virtual surface (i.e., tapping in the air; Tajadura-Jiménez et al., 2015b). Similarly to the previous case, we show that altering the audio feedback related to the level of the tapping strength changes both the tapping behavior and the perceived hardness of the surface. Importantly, results obtained by using measures such as

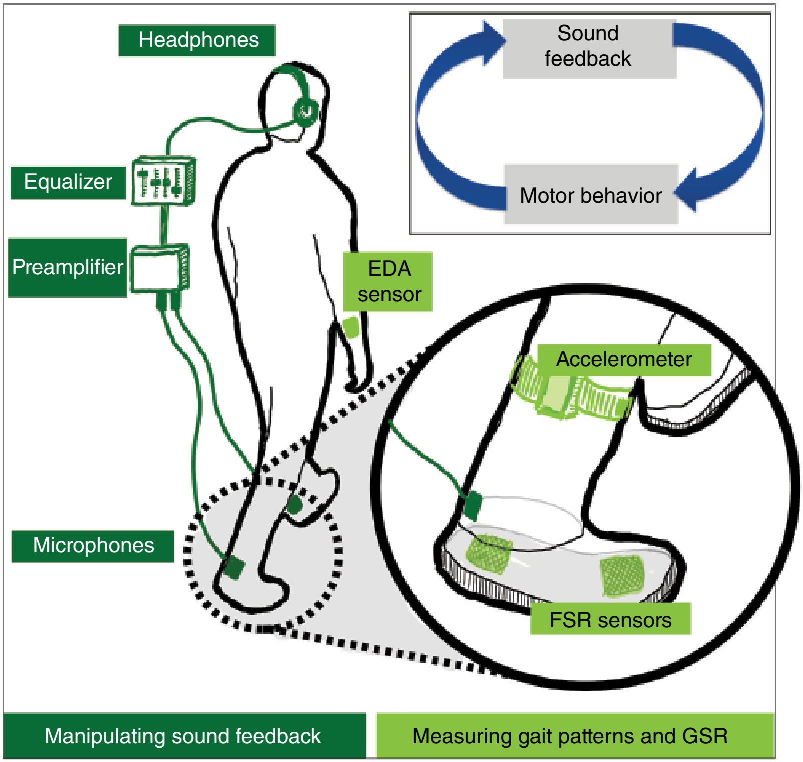


Figure 18.2 Bodily interactions with objects (e.g., walking on a ground surface) produce sound. The sound varies depending on the object material (e.g., ground material) and on the interaction (motor) behavior (in this case gait patterns), which continuously "sculpts" the feedback sound. The sound feedback is used to adjust the motor behavior, thus closing the action-perception loop shown in the subpanel on the top—right corner. The system displayed was used to alter in real time the sound feedback resulting from walking, and to measure the effects in estimated body size, gait patterns (measured with accelerometer and force sensitive resistors—FSR), and in arousal (measured by an electrodermal activity sensor—EDA). *Source:* Figure adapted from Tajadura-Jiménez et al. (2015a). © 2015 ACM, Inc. Retrieved from https://doi.org/10.1145/2702123.2702374. Reprinted by permission.

the electrodermal activity (EDA) and emotion questionnaires showed that the emotional state is also affected and should be taken into account. Summing the different studies indicates that the modelling of such a perception-action loop between touch and sound is rather complex. Both the expectations we have about the experience of touching an object and the perception of the material as we touch the object influence *how* we touch it (Bianchi-Berthouze & Tajadura-Jiménez, 2014).

Initial Learning from Work in Physical Rehabilitation

In this section we discuss applications of the opportunities offered by the body-centered sound mechanisms discussed in the previous sections. Whereas from the previous sections, we have already acquired a flavor of the variety of areas of application, we focus here on a specific context of applications, that is, physical rehabilitation. Our selection is motivated by the fact that physical rehabilitation requires addressing many of the issues discussed above, that is, enhancing and remapping body perceptions and body capabilities as well as related emotional needs. We will review some specific work done in this field of physical rehabilitation discussing the opportunities, complexities, and challenges brought by the use of a body-centered sound design and ubiquitous technologies, and the new questions and possibilities these technologies open to strengthen applications in many other fields.

Sound as physical information

Movement sonification (i.e., real-time auditory feedback of motion) is being increasingly used in the context of physical rehabilitation. Applications have emerged, for example, for recovery of movement postoperation, for reacquiring lost motor capabilities (Vogt et al., 2009) due, for example, to stroke (Wallis et al., 2007), or to facilitate movement despite pain and fear of injury (Singh et al., 2015). Sound is used here to enhance a feedback that is generally not visible or audible—proprioceptive (kinesthetic) feedback. One of the main reasons for using sound as a form of feedback in physical rehabilitation is that it has been shown to facilitate sensory-motor learning (see above). In addition, sound can be played through portable speakers and/or headphones, and as it does not require fixating a display, it is suitable for any type of exercise and movement independently of the body configuration. This is even more important nowadays as physical rehabilitation is moving away from simple physical exercise dedicated sessions into functional activity. Hence, auditory feedback provided through wearable mobile devices becomes a suitable opportunity compared to visual feedback (Singh et al., 2015). Despite these interesting properties, extended experiments on movement sonification or wide clinical trials are still sparse.

Auditory feedback to guide movement The initial use of sonification during physical rehabilitation aimed to inform the patients or the clinician that the movement deviated from a set trajectory. Maulucci and Eckhouse (2000) successfully showed that this use of augmented auditory feedback was beneficial for a rehabilitation task (in comparison to a group receiving no auditory feedback). In their work, no sound was heard when the movement trajectory of the tracked limb was kept within a defined margin, whilst changes in sound frequency indicated the amount of deviation of the movement from the ideal trajectory. In this case the auditory feedback can be considered as a gradual alarm whose sound characteristics informed about the importance of the problem and the need to attend to it (this type of feedback is generally formalized as knowledge of results—KR).

Given the positive effect on motor learning, the use of sound feedback has been increasingly used to guide movement and enhance the perception of moving. Alterations of the sound are used to maintain movement within a desired movement trajectory and kinematic profile. For example, an increase in speed may be signaled by an increase in sound volume, whereas a deviation from a set direction may lead to a distorted sound or modified pitch. Using such a sonification, Boyer, Pyanet, Hanneton, and Bevilacqua (2014) showed that users could follow a specific velocity profile of the arm movement. Interestingly, most users were able to adapt to velocity profiles that changed over time, without being conscious of such shifts. Such a use of movement sonification could inform and help a person following a movement path and dynamics while discouraging that person from using compensatory movements, hence leading to a more effective therapy.

In the cases described above, other than guiding movement trajectory and its kinematics, sonification is used to enhance perception of target reaching. This is either indicated with a specific change in sound or by the ending of the sound. The sonification of reaching a target is considered very important as it provides motivation to cope with the challenging activity by having a clear and rewarding ending point (goal). Rewarding set targets has been shown to lead to increases in performance (e.g., Newbold, Bianchi-Berthouze, Gold, Tajadura-Jimémez, & Williams, 2016; Singh et al., 2015; Wallis et al., 2007) and also to an increased sense of achievement and self-efficacy (Singh et al., 2014) even when performance did not change.

Calibration of movement path, acceptable deviation range and targets are set and adjusted by taking into account the physical capabilities of the person. As in the case of visual-based feedback systems (Lam, Varona-Marin, Li, Fergenbaum, & Kulić, 2016), the setting of these parameters is generally performed by clinical staff. However, as physical rehabilitation moves into self-management, researchers are investigating methods to facilitate either automatic calibration or a calibration that allows patients' exploration of their physical capabilities. An example is discussed at the end of this section.

Choice of parameters for movement-sound mapping A question that is still open is the type of sounds that should be used to map movement (Dubus & Bresin, 2013). Indeed, there are several possible approaches to design movement-sound mappings. To be promptly used, even if the sound is simple, the information it carries needs to be intuitive (i.e., easily learned). Researchers have investigated how sound characteristics relate to movement characteristics. Following the review of related work by Bresin and Friberg (2013) showing that generally, vertical movement is more easily associated with increase in pitch, Scholz et al. (2014) investigated the perception of tone pitch and tone brightness in relation to vertical and horizontal movement in the elderly. They found that the mapping of changes in pitch could be learned easily both in relation to horizontal and vertical movement, whereas brightness was more easily mapped to horizontal movement. These results show the importance of carefully selecting the sonification mapping and sound parameters for the mapping to be intuitive in demanding and often anxiety-triggering task such as those performed in physiotherapy.

In summary, different approaches reside in either sonifying the errors between the movement and "ideal" trajectory or sonifying the movement itself; the difference in effects between these two approaches remains unclear and asks to be explored widely. In this latter approach, the auditory feedback can be considered as enhancing body awareness, which could lead to a faster sensory-motor learning. Important questions remains on the different involved mechanisms: is the learning mechanism effective by enhancing the proprioception through sound or is the alteration of the expected sound that produced the movement correction? In all cases, the emotional response (beyond motivation and engagement) to the action-perception loop has been often eluded, but it seems important to consider it, even when using simple sounds, as discussed in the next section.

Finally, sonification can support auditory-motor integration and hence does not only provide a form of sensory supplement when the proprioceptive system is dysfunctional but, as suggested by Scholz et al. (2014), it also offers a way to establish brain networks that transform sound into movement (e.g., Altenmüller, Marco-Pallares, Münte, & Schneider, 2009). Instead, the alarm-like approach may be considered more useful during functioning where too much information would be overwhelming or not socially acceptable (Singh, Bianchi-Berthouze, & Williams, 2017).

Sound to address emotional barriers

Whilst the above sonification frameworks aimed to compensate for limited proprioceptive feedback, to provide information about the quality of movement, or to promote correct movement and a more engaging activity, emotional factors are also critical barriers to physical rehabilitation. Low engagement, low self-efficacy, low self-esteem, pain, and fear of injury are some of the factors that affect the efficacy of physical rehabilitation and often lead to low adherence. Unfortunately, emotional aspects are often ignored in the design of physical rehabilitation technology.

Using musical and naturalistic sounds as metaphors A particular approach aimed to motivation has explored the use of music instead of simple sonifications. Huang et al. (2005) created a virtual environment with both visual and auditory feedback, and evaluated it with patients. They used general musical concepts that could potentially favor the user engagement (Chen et al., 2006). Further, the *Musical Stroke project* (Kirk, 2015; see also https://strokeproject.wordpress.com), with a strong user-centered design, focused in altering the sound feedback received during interaction with objects as a way to transform everyday objects in musical instruments. In this way, they could engage people who suffered stroke with their physical rehabilitation and enhance their motivation to perform the exercises they have been recommended to do. This work links to our discussion of the possibilities of changing interactions with objects via sound as a way to enhance movement dynamics, which, as exemplified here, has applicability in the context of physical rehabilitation.

To make the sonification more intuitive and inviting, some researchers have also explored the use of metaphors (see for instance, Rasamimanana et al., 2011; Robby-Brami et al., 2014; Bevilacqua & Schnell, 2016). For example Vogt et al. (2009), make use of naturalistic sounds to sonify the movement during shoulder physiotherapy. As the arm is moved upwards, the sonification changes from woodland sounds, to animals sounds, to the winds sound through the trees. The aim was to provide a more clear (even if gross) understanding of the position of the arm and, at the same time, a more aesthetically pleasant and relaxing experience.

The use of naturalistic sound was also investigated in Singh et al. (2015), where the phases of a movement are mapped into a path in a naturalistic environment (e.g., a forest) and changes in sounds reflect the reaching of a new landmark (e.g., a river, a herd of sheep). The skipping of one type of sound indicated that a phase of a movement had been avoided (e.g., the bending forward of the trunk during a sit-to-stand movement). The aim was again to enhance the sense of moving, to provide broad information of the movement quality, and also to relax. Singh et al. (2015) also showed that naturalistic sounds were relaxing, and possibly more enjoyable; however, when sonification was used to understand one's body movement capability during pain or anxiety-triggering movement, the mapping between sound and movement aspects needed to be simple, i.e., easy to understand. Their participants showed preference towards a simple sequence of tone scales (as explained in the next subsection) rather than naturalistic sound or complex music, as they could better interpret the progression of their movement and the reaching of milestones in the simple sequence of tone scales.

Sonification to enhance the feeling of being capable (perceived self-efficacy) and confidence The mechanisms discussed above show that auditory feedback could be used to alter one's perception of body capabilities with possible emotion regulation effect and behavior changes. Singh et al. (2014, 2015) explored this possibility. They investigated how movement sonification feedback should be designed to address various psychological barriers to physical activity rather than just the physical ones and motivation. In the context of physical rehabilitation of chronic pain, they have proposed and evaluated a sonification framework called *Go-with-the-Flow* to address perceived self-efficacy, confidence in moving, and fear of pain.

People with chronic pain often develop fear or anxiety towards movement that is perceived as threatening (e.g., stretching; Turk & Rudy, 1987). Rather than calibrating the sonification parameters to the physical capability of a person and to the expected movement profile, in the *Go-with-the-Flow* framework the setting of the parameters reflects the psychological capability of a person in performing a movement. The *Go-with-the-Flow* framework proposes the use of discrete tone scale to sonify progress of movement during the execution of anxiety inducing exercise.

As fear and lack of confidence in one's capabilities may deter people from engaging in a movement, the *Go-with-the-Flow* framework proposes to sonify each phase of a movement according to its psychological demand. Rather than using biomechanical phases of a movement, the patient is asked to explore a movement and to split the movement into phases. For example, physiotherapists and people with chronic pain identify three phases of a stretching forward movement: the first phase corresponds to the initial part of the stretching to the point where the person stops feeling comfortable stretching forward. The second phase is more challenging as anxiety kicks in even if the stretching is within a target set by the person. A third phase is the one beyond the target, i.e., where the person feels that the stretching is beyond his / her capabilities or state.

In their work, Singh et al. (2014, 2015) showed that when using an increasing tone scale to sonify the first phase, people reported the feeling of having climbed a mountain, and that when they reach the higher tone they feel a sense of achievement and they want to go forward (see Figure 18.3). The increasing tone scale was used to provide a sense of increased challenge knowing that this phase of the movement was considered feasible despite the pain. The sonification of the second phase takes advantage of the emotional change and increasing coping capabilities created by the first sonification, and uses a decreasing tone scale to provide a sense of arrival and conclusion.

Finally, in order to further bust self-efficacy, the self-calibration aims to set the amount of reward to provide according to the set amount of stretching. Upon calibration (i.e., the setting of the maximum amount of stretching the person feels capable of), the step between the playing of two consecutive tones is automatically set on the basis of how much the person feels capable of doing. For example, people who can stretch their trunks only for few degrees will be rewarded more frequently (e.g., at every degree of trunk movement as they bend forward) than people that can stretch

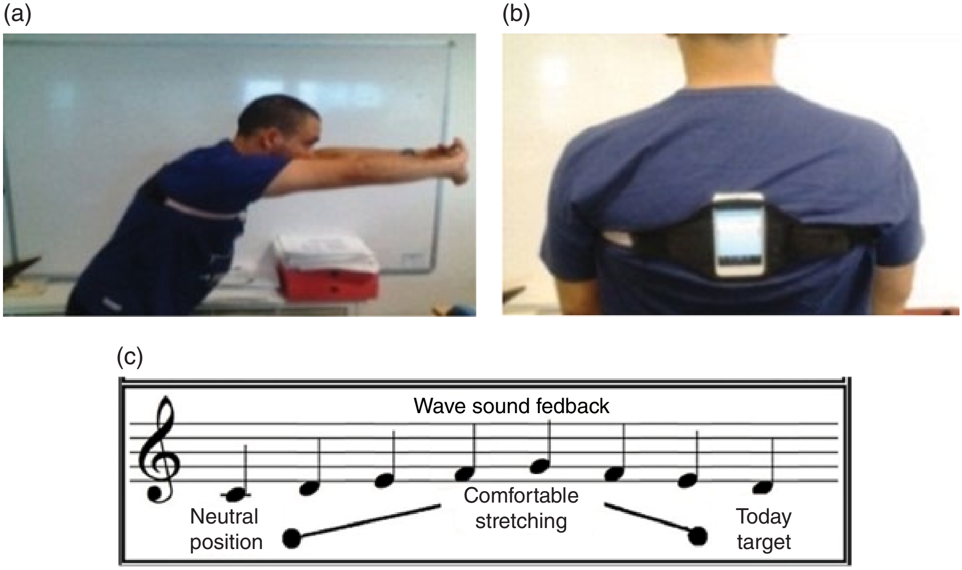


Figure 18.3 (a) and (b) A smartphone attached to the back of the person is used to track the movement of the trunk and provide real-time feedback. (c) Sonification used to facilitate the stretching forward movement *Source:* Singh et al. (2014, 2015).

longer. In this way, despite the different capabilities, their stretch forward movement will be mapped over the same number of progressive tones. The faster rewards aim to enhance very restricted movement, producing a sense of "*worth doing and being capable*" (as participants reported). The slower rewards for the less restricted movement aim to provide a more clear sense of progress. In turn, this brings a sense of control as the body travels a longer way and it is more challenged.

Control studies showed the benefits of such a structured and self-calibrated sonification in terms of increased self-efficacy, awareness, and even increase in physical performances. Participants also reported that the simple sequence of discrete-scale tone-based sonification was enjoyable despite its simplicity and, most importantly, it was rich in information and easy to map during a psychologically demanding movement.

Sensory substitution: using sonification structure to overwrite proprioceptive feedback Discussion with physiotherapists and people with chronic pain raised the question of how to sonify movement targets. Two important issues emerged: how to avoid overdoing (stopping at set targets), and at the same time how to encourage progress when the physical capabilities are there. Building on the *Go-with-the-Flow* framework, and taking advantage of the embodiment of sound, Newbold et al. (2016) proposed to use musically informed sonification to provide different levels of closure of the phrases of sonification. Using the concept of cadence, they investigated if a stable cadence would lead the body to naturally slow down as it approaches the target, whereas an unstable cadence would lead the body to want to continue to move passed the target (see Figure 18.4). They tested the different levels of stability, and they show that stable cadences let people stop soon after the end of the sonification,

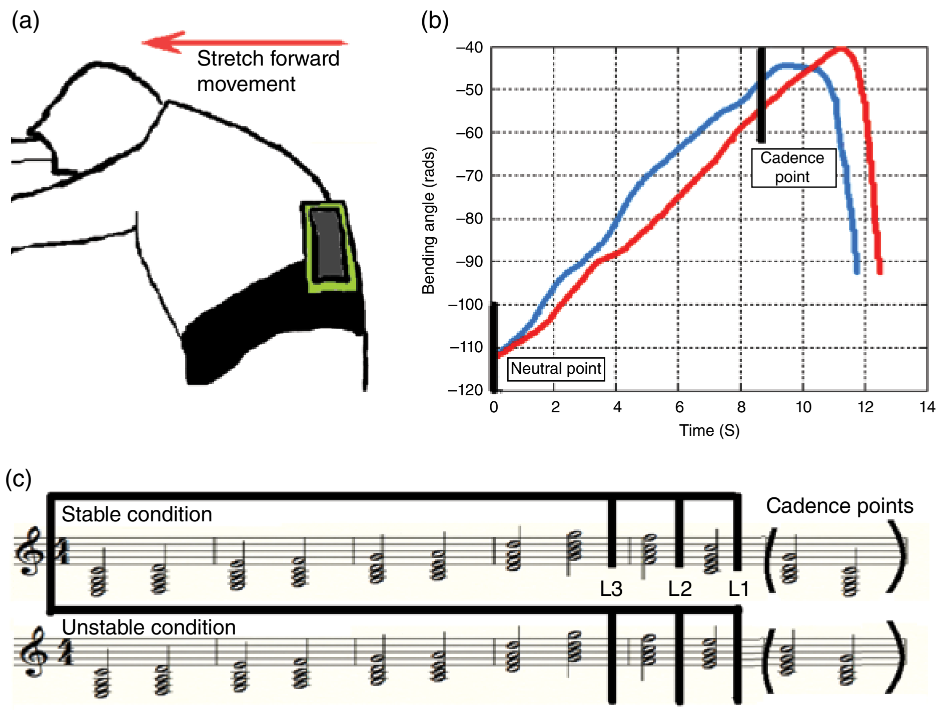


Figure 18.4 (a) shows the stretch forward movement (trunk bend angle) tracked by a smartphone-based system. (b) An example of the output recorded for one participant: the movement towards and past the cadence point for stable (in blue/black) and unstable (red/grey) conditions. This example illustrates the difference in amount of stretching beyond the cadence point. (c) An overview of the sonification stimuli used, all ending on either a stable or unstable cadence point, and the three lengths L1, L2 and L3 (derived by removing one or two chords from before the cadence point as shown, followed by either the stable or unstable cadence). *Source:* Figure adapted from Newbold et al. (2016) © 2016 ACM, Inc. Retrieved from https://doi.org/10.1145/2858036.2858302. Reprinted by permission.

whereas with unstable cadence, participants continued stretching forward for longer when the sound stopped. In addition, the change in speed profile was more sudden with unstable cadences. Self-report questionnaires also show that unstable cadences provided a sense of wanting to continue to move compared to stable cadences.

Whilst music-based sonification and simple sonification have been shown to be effective and play different roles in supporting people during physical rehabilitation, an important concept emerges with respect to overcoming emotion. One could argue that, in order for the feedback to be effective, the sonification approach used needs to enhance the sense of agency so that the information is processed in a bottom-up manner and perceived as direct-body feedback. This is particularly critical when the sonification itself aims not just to inform but to regulate critical emotional states. So this possibly raises future research questions such as: what does the brain perceives as directly created by one's body? Or, what type of exposure does one require for the sonification to be perceived as a direct production of one's body action?

Conclusions and Future Directions

In this chapter we have summarized some of the mechanisms identified in neuroscience and psychology research. These mechanisms can serve as bases for better understanding users' experience of interactive technologies. Moreover, they can also serve as principles to design more effective sound feedback for these applications by adopting a body-centered approach. Below we summarize some of these principles and takeaway messages.

Summary topics important for future applications

Sonic self-avatar as representation of oneself As shown above, sounds constituting our sonic body can be instrumental for creating new experiences and shaping the mental representations of our body. A number of parameters, however, should be taken into account when creating a sonic-self avatar. These include the context (e.g., engine sound in the driving simulation), latency of the feedback (e.g., delay between the tapping action and the tapping sound in the body extension experiments), spatial location (e.g., the location of the heartbeat sound sources), and congruency with other sensory information (e.g., vibrations accompanying the engine sound).

Sensory enhancement and augmentation As we showed, sound is often used to sonify the invisible—like hearing the body movement dynamics and its success in rehabilitation or sports, to amplify the feeling of one's own heart beating, to bring back the lost sense of touch in prosthetic applications, or to make one's footsteps really noticeable and instrumental to change one's body representation and gate. Such auditory-based augmentation of our reality can only work in terms of provided sensory enhancement if conditions of time, space, intensity and context are respected (see next point). Cognitive load and attention demands should also be taken into account as a limiting factor of such sensory augmentation (e.g., Vazquez-Alvarez & Brewster, 2011).

Respecting thresholds in action—perception loops It is not enough to sonify the user's movement and actions but it is important to do so in such a way that the user can intuitively incorporate the sound feedback provided into the existing action-perception loop. For example, in order for the sound feedback to be considered as being generated by one's body (i.e., feeling of agency) and therefore to have the possibility to produce unconscious adjustments in movement or in body representations, certain thresholds of deviations from the predicted sound feedback should not be trespassed. Thresholds apply to all temporal, spatial, and intensity deviations. Specifically, if sound feedback is presented in large asynchrony with respect to the instant in which the action was generated, or if the sound feedback is presented too far away from one's body (e.g., outside the PPS), or if deviations in acoustic parameters (e.g., intensity of the sound feedback) are too large from the sound predicted based on the performed action, then sounds will no longer be perceived as produced by one's body, as found in the examples with tapping action and tapping sounds mentioned above.

Emotional design Another take-away message when designing applications is the necessity to think about sound design in a unifying way that joins physical and psychological aspects of the user experience, like in the case of rehabilitation.

Emotional design brings, on the one hand, the necessity to consider user's emotional state (e.g., anxiety) and how it influences attention, cognitive, and perceptual processes, which will affect how people perceive and react to the presented sound feedback. On the other hand, it opens possibilities of using sound feedback in a way that triggers specific intuitive responses in specific situations (e.g., alerts), relax the user in other situations, or interact with other processes in modulating body representations. Sound design should consider the choice of parameters (i.e., intensity, source location, associated meaning) in relation to the emotional responses it can induce according to the context of use and users' individual psychological and physical circumstances.

Future outlook

The exponentially growing use of wearable sensors and computing, including various mobile, augmented reality devices like augmented glass, wristbands, and smart fabrics, which allow for monitoring body motion and other physiological states of the body, inevitably invites a search for ways to convey such bodily information to the user. We briefly outline a number of potential applications that consider sound and body:

Sounding body/sounding suit A potential application would be one that offers the possibility of enhancing people's awareness of their body movement, body parts location, and size by augmenting the sounds, or presenting new sounds, which the body produces when moving. A previous study has shown the potential of delivering intensive somatosensory stimulation to the whole body, by asking people to wear a custom-made neoprene diving suit, to improve body image in treatments of anorexia nervosa (Grunwald & Weiss, 2005). Another study has shown that people may feel their body as being metallic or "robotized" if specifically designed auditory and vibrotactile feedback accompanies the flexing of the joins (Kurihara, Hachisu, Kuchenbecker, & Kajimoto, 2013). Would it be possible to increase the positive awareness of one's body through a specifically designed sounding suit? And could such suit be used to enhance the wellbeing of the general population and the population with clinical conditions, such as anorexia nervosa or obesity?

Having a sound system that raises awareness of one's everyday movement and aims at reeducating movement habits can have a strong therapeutic impact (Wang, Turaga, Coleman, & Ingalls, 2014; see also Schiphorst, 2011). Indeed, the area of "somatic movement education and therapy" is closer to therapy than to classical dance (Eddy, 2009). The question, here, is the movement-to-sound mapping strategy and selection of appropriate spatio-temporal relationships for selecting the important body motion features.

Sonifying quantified self An emerging movement of quantified self deals with person's daily life in terms of inputs (e.g., air, food) and one's performance (Russo, 2015). Recent advances in wearable computing, such as sensor arrays for real-time analysis of body chemistry (Gao et al., 2016), provide users with various multichannel data streams that are inherently related to our body and can fully benefit from body-centered sound design. In other words, how does your glucose level sound today? It would be interesting to see how such real-time audio-chemical loops may influence one's organism.

Body extension/body swap Our experiments are showing some apparent limits to the plasticity of one's body representation. These limits might be related to the special processing of auditory information in the PPS (as opposed to the far space), or perhaps they have to do with one's body model not being totally plastic. For instance, there are much fewer reports in previous literature of sensory-induced body shrinkage illusions than of body expansion illusions (although see, for instance, Banakou, Groten, & Slater, 2013; van der Hoort et al., 2011). Nevertheless, perhaps the observed limits have to do with the brief exposure to auditory feedback during the previously conducted studies. Would it be possible to further extend or shrink the body, or to have the experience of having a body of a shape and appearance very different to our actual body?

Sensory enhancement and magical experiences Apart from benefits described above, sound rendering technologies often offer an interesting solution like bone-conducted sound or narrow sound beams for personal experience using ultrasound loudspeakers (Olszewski, 2009). Such technologies can be used for creating unique or, as Slater and Usoh (1994) term them, magical experiences. For example, a simulation of hearing one's inner voice can be used as sonic avatar-based therapy for persecutory auditory hallucinations (Leff, Williams, Huckvale, Arbuthnot, & Leff, 2014). The important question here is what level of technological simulation transparency needs to be maintained to keep the "the illusion of nonmediation" (Lombard & Ditton, 1997).

Sensory substitution systems Sound has been used to provide missing visual or tactile input (e.g., Väljamäe & Kleiner, 2006). Both vibrotactile and auditory cues are used to indicate the location of objects approaching one's body in the absence of vision (e.g., Auvray, Hanneton, & O'Regan, 2007; Bird, Marshall, & Rogers, 2009). In terms of enhancing proprioception via sound, people with a lost sense of touch use an audio glove (Lanzetta et al., 2004). It remains to be answered whether new brain networks are established when sound is used to provide a form of sensory supplement when the proprioceptive system is dysfunctional.

Acknowledgments

Ana Tajadura-Jiménez acknowledges support of the ESRC ES/K001477/1 "The hearing body" grant, the MINECO Ramón y Cajal research contract RYC-2014-15421 and the MINECO grant PSI2016-79004-R (AEI/FEDER, UE). Frederic Bevilacqua acknowledges support of the Laboratory of Excellence SMART (ANR-11-LABX-65). Nadia Bianchi-Berthouze acknowledges support of the EPSRC EP/H017178/1 "Emo&Pain" grant.

References

Aglioti, S. M., & Pazzaglia, M. (2010). Representing actions through their sound. *Experimental Brain Research, 206*(2), 141–151.

Albers, S. (2003). *Eating mindfully: How to end mindless eating and enjoy a balanced relationship with food*. Oakland, CA.: New Harbinger Publications.

Altavilla, A., Caramiaux, B., & Tanaka, A. (2013). Towards gestural sonic affordances. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (NIME'13) (pp. 61–64). Berlin: Springer.

Altenmüller, E., Marco-Pallares, J., Münte, T. F., & Schneider, S. (2009). Neural reorganization underlies improvement in stroke-induced motor dysfunction by music-supported therapy. *Annals of the New York. Academy of Science 1169*, 395–405.

Asutay, E., & Västfjäll, D. (2015). Attentional and emotional prioritization of the sounds occurring outside the visual field. *Emotion, 15*(3), 281–286.

Auvray, M., Hanneton, S., & O'Regan, K. (2007). Learning to perceive with a visuo-auditory substitution system: localization and object recognition with the voice. *Perception, 36*(3), 416–430.

Banakou, D., Groten, R., & Slater, M. (2013). Illusory ownership of a virtual child body causes overestimation of object sizes and implicit attitude changes. *Proceedings of the National Academy of Sciences, 110*(31), 12846–12851.

Bevilacqua, F., Boyer, E. O., Francoise, J., Houix, O., Susini, P., Roby-Brami, A., & Hanneton, S. (2016). Sensori-motor learning with movement sonification: A perspective from recent interdisciplinary studies. *Frontiers in Neuroscience, 10*: 385. doi:10.3389/fnins.2016.00385

Bevilacqua, F., & Schnell, N. (2016). From musical interfaces to musical interactions. In A. Gaggioli, A. Ferscha, G. Riva, S. Dunne, & I. Viaud-Delmon (Eds.), *Human computer confluence: Transforming human experience through symbiotic technologies* (pp. 125–140). Warsaw: De Gruyter Open Ltd. doi:10.1515/9783110471137-007

Bevilacqua, F., Schnell, N., Fdili Alaoui, S., Klein, G., & Noeth, S. (2011). Gesture capture: Paradigms in interactive music/dance systems. *Emerging Bodies: The Performance of Worldmaking in Dance and Choreography, 183.*, 183–193.

Bianchi-Berthouze, N., & Tajadura-Jiménez, A. (2014, April/May). *It's not just what we touch but also how we touch it.* Paper presented at the TouchME workshop, CHI Conference on Human Factors in Computing Systems, Toronto.

Bird, J., Marshall, P., & Rogers, Y. (2009). Low-fi skin vision: a case study in rapid prototyping a sensory substitution system. In *Proceedings of the 23rd British HCI Group Annual Conference on People and Computers: Celebrating People and Technology* (pp. 55–64). London: British Computer Society.

Blanke, O., Slater, M., & Serino, A. (2015). Behavioral, neural, and computational principles of bodily self-consciousness. *Neuron, 88*(1), 145–166.

Botvinick, M., & Cohen, J. (1998). Rubber hands "feel" touch that eyes see. *Nature, 391*(6669), 756.

Boyer, E. (2015). Continuous auditory feedback for sensorimotor learning (Unpublished PhD thesis). Université Pierre et Marie Curie, Paris.

Boyer, E., Pyanet, Q., Hanneton, S., & Bevilacqua, F. (2014). Learning movement kinematics with a targeted sound. *Lecture Notes in Computer Science, 8905*, 218–233.

Boyer, E. O., Vandervoorde, L., Bevilacqua, F., & Hanneton, S. (2015). Touching sounds: Perception of the curvature of auditory virtual surfaces. Proceedings of 2015 IEEE Virtual Reality (VR) (pp. 153–154). doi:10.1109/VR.2015.7223341

Bradley, M. M., & Lang, P. J. (1999). *International affective digitized sounds (IADS): Stimuli, instruction manual and affective ratings.* Gainesville, FL: The Center for Research in Psychophysiology, University of Florida.

Bresin, R., de Witt, A., Papetti, S., Civolani, M., & Fontana, F. (2010, April). *Expressive sonification of footstep sounds.* Paper presented at the Interaction Sonification workshop (ISon) Stockholm, Sweden.

Bresin, R., & Friberg, A. (2013). Evaluation of computer systems for expressive music performance. In Kirke, A. & Miranda, E. R. (Eds.), *Guide to computing for expressive music performance* (pp. 181–203). London: Springer.

Brozzoli, C., Gentile, G., Petkova, V. I., & Ehrsson, H. H. (2011). fMRI adaptation reveals a cortical mechanism for the coding of space near the hand. *The Journal of Neuroscience, 31*, 9023–9031.

Cabe, P. A., & Pittenger, J. B. (2000). Human sensitivity to acoustic information from vessel filling. *Journal of Experimental Psychology: Human Perception and Performance, 26*(1), 313–324.

Caclin, A., Soto-Faraco, S., Kingstone, A., & Spence, C. (2002). Tactile "capture" of audition. *Perception and Psychophysics, 64*(4), 616–630.

Caramiaux, B., Altavilla, A., Pobiner, S. G., & Tanaka, A. (2015). Form follows sound: Designing interactions from sonic memories. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, CHI '15* (pp. 3943–3952). New York, NY: ACM.

Caramiaux, B., Schnell, N., Françoise, J., & Bevilacqua, F. (2014). Mapping through listening. *Computer Music Journal, 38*, 34–48.

Carney, D. R., Cuddy, A. J., and Yap, A. J. (2010). Power posing brief nonverbal displays affect neuroendocrine levels and risk tolerance. *Psychological Science 21*, 1363–1368.

Carruthers, G. (2008). Types of body representation and the sense of embodiment. *Consciousness and Cognition, 17*(4), 1302–1316.

Castiello, U., Giordano, B. L., Begliomini, C., Ansuini, C., & Grassi, M. (2010). When ears drive hands: The influence of contact sound on reaching to grasp. *PLoS ONE, 5*(8), e12240.

Chen, Y., Huang, H., Xu, W., Wallis, R. I., Sundaram, H., Rikakis, T., … & He, J. (2006, October). The design of a real-time, multimodal biofeedback system for stroke patient rehabilitation. In *Proceedings of the 14th ACM International Conference on Multimedia* (pp. 763–772). doi:10.1145/1180639.1180804

Chieffi, S., Fogassi, L., Gallese, V., & Gentilucci, M. (1992). Prehension movement directed to approaching objects: influence of stimulus velocity on the transport and the grasp components. *Neuropsychologia, 30*, 877–897.

Csapó, Á., Wersényi, G., Nagy, H., & Stockman, T. (2015). A survey of assistive technologies and applications for blind users on mobile platforms: A review and foundation for research. *Journal on Multimodal User Interfaces, 9*(4), 275–286.

Damasio, A. R. (1999). *The feeling of what happens: body and emotion in the making of consciousness* (1st ed.). New York, NY: Harcourt Brace.

Damasio, A., & Damasio, H. (2006). Minding the body. *Daedalus, 135*(3), 15–22.

de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition and Emotion, 14*(3), 289–311.

Delaherche, E., Chetouani, M., Mahdhaoui, A., Saint-Georges, C., Viaux, S., & Cohen, D. (2012). Interpersonal synchrony: A survey of evaluation methods across disciplines. *IEEE Transactions on Affective Computing, 3*(3), 349–365.

della Volpe, R., Popa, T., Ginanneschi, F., Spidalieri, R., Mazzocchio, R., & Rossi, A. (2006). Changes in coordination of postural control during dynamic stance in chronic low back pain patients. *Gait and Posture, 24*(3), 349–355.

Deutschmann, M., Steinvall, A., & Lagerström, A. (2011). Gender-bending in virtual space: Using voice-morphing in Second Life to raise sociolinguistic gender awareness. In *Proceedings of the Learning a Language in Virtual Worlds : A Review of Innovation and ICT in Language Teaching Methodology (V-lang) International Conference* (pp. 54–61). Warsaw: Warsaw Academy of Computer Science, Management and Administration.

de Vignemont, F. (2010). Body schema and body image—Pros and cons. *Neuropsychologia, 48*(3), 669–680.

de Vignemont, F., Ehrsson, H. H., & Haggard, P. (2005). Bodily illusions modulate tactile perception. *Current Biology, 15*(14), 1286–1290.

Dodds, T. J., Mohler, B. J., & Bülthoff, H. H. (2011). Talk to the virtual hands: Self-animated avatars improve communication in head-mounted display virtual environments. *PLoS One 6*(10), e25759.

Dubus, G. & Bresin, R. (2013). A systematic review of mapping strategies for the sonification of physical quantities. *PLoS ONE 8*(12): e82491.

Eddy, M. (2009). A brief history of somatic practices and dance: Historical development of the field of somatic education and its relationship to dance. *Journal of Dance and Somatic Practices, 1*(1), 5–27.

Edworthy, J. (2013). Medical audible alarms: A review. *Journal of the American Medical Informatics Association, 20*(3), 584–589.

Ekman, P. (1980). The facial features of emotion. *Recherche, 11(*117), 1408–1415.

Ferri, F., Tajadura-Jiménez, A., Väljamäe, A., Vastano, R., & Costantini, M. (2015). Emotion-inducing approaching sounds shape the boundaries of multisensory peripersonal space. *Neuropsychologia, 70*: 468–475.

Françoise, J., Fdili Alaoui, S., Schiphorst, T., & Bevilacqua, F. (2014). Vocalizing dance movement for interactive sonification of Laban effort factors. In *Proceedings of the 2014 Conference on Designing Interactive Systems, DIS '14* (pp. 1079–1082). doi:10.1145/2598510.2598582

Franinovic, K., & Serafin, S. (Eds.) (2013). *Sonic interaction design*. Cambridge, MA: MIT Press.

Gallagher, S. (2005). *How the body shapes the mind*. Oxford: Clarendon Press.

Gao, W., Emaminejad, S., Nyein, H. Y. Y., Challa, S., Chen, K., Peck, A.,… Lien, D. H. (2016). Fully integrated wearable sensor arrays for multiplexed in situ perspiration analysis. *Nature, 529*, 509–514.

Gaver, W. W. (1991). Technology affordances. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 79–84). doi:10.1145/108844.108856

Gaver, W. W. (1993) What in the world do we hear? An ecological approach to auditory event perception. *Ecological Psychology, 5*(1), 1–29.

Gibson, J. J. (1986). *The ecological approach to visual perception*. New York, NY: Routledge.

Giordano, B. L., Egermann, H., & Bresin, R. (2014). The production and perception of emotionally expressive walking sounds: Similarities between musical performance and everyday motor activity. *PLoS ONE, 9*(12), e115587.

Giordano, B. L. & McAdams, S. (2006). Material identification of real impact sounds: Effects of size variation in steel, glass, wood, and plexiglass plates. *Journal of the Acoustical Society of America, 119*(2): 1171–1181.

Giordano, B. L., Visell, Y., Yao, H. Y., Hayward, V., Cooperstock, J. R., & McAdams, S. (2012). Identification of walked-upon materials in auditory, kinesthetic, haptic, and audio-haptic conditions. *Journal of the Acoustical Society of America, 131*, 4002–4012.

Godøy, R. I. (2009). Gestural affordances of musical sound. In R. I. Godøy & M. Leman (Eds.), *Musical gestures: Sound, movement, and meaning* (pp. 103–125). New York: Routledge.

Grassi, M., Pastore, M., & Lemaitre, G. (2013). Looking at the world with your ears: How do we get the size of an object from its sound? *Acta Psychologica, 143*(1), 96–104.

Graziano, M. S., & Cooke, D. F. (2006). Parieto-frontal interactions, personal space, and defensive behavior. *Neuropsychologia, 44*, 2621–2635.

Großhauser, T., Bläsing, B., Spieth, C., & Hermann, T. (2012). Wearable sensor-based real-time sonification of motion and foot pressure in dance teaching and training. *Journal of the Audio Engineering Society, 60*(7/8), 580–589.

Grunwald, M., Weiss, T. (2005). Inducing sensory stimulation in treatment of anorexia nervosa. *QJM, 98*(5), 379–380.

Guest, S., Catmur, C., Lloyd, D., & Spence, C. (2002). Audiotactile interactions in roughness perception. *Experimental Brain Research, 146*(2), 161–171.

Haggard, P., Kitadono, K., Press, C., & Taylor-Clarke, M. (2006). The brain's fingers and hands. *Experimental Brain Research, 172*(1), 94–102.

Hall, C., Pongrac, J., & Buckholz, E. (1985). The measurement of imagery ability. *Human Movement Science, 4*(2), 107–118.

Hayduk, L. A. (1983). Personal space: Where we now stand. *Psycholical Bulletin, 94*, 293–335.
Head, H., & Holmes, H. G. (1911–1912). Sensory disturbances from cerebral lesions. *Brain 34*, 102–254.
Hermann, T., & Ritter, H. (2004). Sound and meaning in auditory data display. *Proceedings of the IEEE, 92*(4), 730–741.
Holmes, N., & Spence, C. (2004). The body schema and multisensory representation(s) of peripersonal space. *Cognitive Processing, 5*(2), 94–105.
Houix, O., Misdariis, N., Susini, P., Bevilacqua, F., & Gutierrez, F. (2014). Sonically augmented artifacts: Design methodology through participatory workshops. In *Proceedings of the 10th International Symposium on Computer Music Modeling and Retrieval (CMMR 2013)* (pp. 20–40). doi:10.1007/978-3-319-12976-1_2.
Huang, H., Ingalls, T., Olson, L., Ganley, K., Rikakis, T., & He, J. (2005). Interactive multimodal biofeedback for task-oriented neural rehabilitation. *Proceedings of the 27th Annual International Conference of the Engineering in Medicine and Biology Society (IEEE-EMBS)* (pp. 2547–2550). doi:10.1109/IEMBS.2005.1616988
James, W. (1890). *The principles of psychology.* New York: Holt.
James, W. (1894) The physical basis of emotion. *Psychological Review, 1*, 516–529.
Jousmäki, V., & Hari, R. (1998). Parchment-skin illusion: sound-biased touch. *Current Biology, 8*(6), R190–R190.
Juslin, P. N., & Västfjäll, D. (2008). Emotional responses to music: The need to consider underlying mechanisms. *Behavioral and Brain Sciences, 31*(5), 559–575; discussion 575–621.
Kenyon, G. P., & Thaut, M. H. (2005). Rhythmic-drive optimization of motor control. In M. H. Thaut (Ed.), *Rhythm, music and the brain: Scientific foundations and clinical applications* (pp. 85–112). New York, NY: Routledge.
Keysers, C. (2010). Mirror neurons. *Current Biology, 19*(21), R971–973.
Kirk, P. (2015). Can specialised electronic musical instruments aid stroke rehabilitation? In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '15)* (pp. 127–132). New York, NY: ACM.
Klatzky, R. L., Lederman, S. J. (2010). Multisensory texture perception. In J. Kaiser & M. J. Naumer (Eds.), *Multisensory object perception in the primate brain* (pp. 211–230). New York, NY: Springer.
Klatzky, R. L., Pai, D. K., & Krotkov, E. P. (2000). Perception of material from contact sounds. *Presence: Teleoperators and Virtual Environments, 9*(4), 399–410.
Kleinsmith, A., & Bianchi-Berthouze, N. (2013). Affective body expression perception and recognition: A survey. *IEEE Transactions on Affective Computing, 4*(1), 15–33.
Kurihara, Y., Hachisu, T., Kuchenbecker, K. J., & Kajimoto, H. (2013). Jointonation: Robotization of the human body by vibrotactile feedback. *Proceedings of the ACM special interest group on Computer GRAPHics and Interactive Techniques (SIGGRAPH)* (pp. 11:1–11:3). New York, NY: ACM.
Lam, A. W. K., Varona-Marin, D., Li, Y., Fergenbaum, M., & Kulić, D. (2016). Automated rehabilitation system: Movement measurement and feedback for patients and physiotherapists in the rehabilitation clinic. *Human–Computer Interaction, 31*, 294–334.
Lanzetta, M., Perani, D., Anchisi, D., Rosén, B., Danna, M., Scifo, P.,…Lundborg, G. (2004). Early use of artificial sensibility in hand transplantation. *Scandinavian Journal of Plastic and Reconstructive Surgery and Hand Surgery, 38*(2), 106–111.
Larsson, P. (2005). *Virtually hearing, seeing, and being: Room acoustics, presence, and audiovisual environments* (Unpublished doctoral dissertation). Chalmers University of Technology, Gothenburg, Sweden.
Larsson, P., Väljamäe, A., Västfjäll, D., Tajadura-Jiménez, A., & Kleiner, M. (2010). Auditory-induced presence in mixed reality environments and related technology. In E. Dubois, P. Gray, & L. Nigay (Eds.), *The Engineering of Mixed Reality Systems* (pp. 143–163). London: Springer.

LeDoux, J. E. (1998). *The emotional brain: The mysterious underpinnings of emotional life.* New York, NY: Simon & Schuster.

Leff, J., Williams, G., Huckvale, M., Arbuthnot, M., & Leff, A. P. (2014). Avatar therapy for persecutory auditory hallucinations: What is it and how does it work? *Psychosis, 6*(2), 166–176.

Lemaître, G., & Heller, L. M. (2013). Evidence for a basic level in a taxonomy of everyday action sounds. *Experimental Brain Research, 226*(2): 253–264.

Lemaître, G., Heller, L. M., Navolio, N., & Zúñiga-Peñaranda, N. (2015). Priming gestures with sounds. *PloS One, 10*(11), e0141791.

Leman, M. (2007). *Embodied music cognition and mediation technology.* Cambridge, MA: MIT Press.

Leman, M., Moelants, D., Varewyck, M., Styns, F., van Noorden, L., & Martens, J. P. (2013). Activating and relaxing music entrains the speed of beat synchronized walking. *PloS one 8*, 7, e67932.

Lenti Boero, D., & Bottoni, L. (2008). Why we experience musical emotions: Intrinsic musicality in an evolutionary perspective. *Behavioral and Brain Sciences, 31*(05), 585–586.

Levenson, R. W. (1994). Human emotions: A functional view. In P. Ekman & R. J. Davidson (Eds.), *The nature of emotion: Fundamental questions* (pp. 123–126). New York, NY: Oxford University Press.

Lewis, J. S., & McCabe, C. S. (2010). Body perception disturbance (BPD) in CRPS. *Practical Pain Management 10*, 60–66.

Lewkowicz, D. J. (1996). Perception of auditory-visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception and Performance, 22*(5), 1094–1006.

Lewkowicz, D. J. (1999). The development of temporal and spatial intermodal perception. In G. Aschersleben, T. Bachman, & J. Musseler (Eds.), *Cognitive contributions to the perception of spatial and temporal events* (pp. 395–420). Amsterdam: Elsevier.

Li, X., Logan, R. J., & Pastore, R. E. (1991). Perception of acoustic source characteristics: Walking sounds. *The Journal of the Acoustical Society of America, 90*(6), 3036–3049.

Linkenauger, S. A., Witt, J. K., & Proffitt, D. R. (2011). Taking a hands-on approach: Apparent grasping ability scales the perception of object size. *Journal of Experimental Psychology—Human Perception and Performance, 37*(5), 1432–1441.

Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, *3*(2). doi:10.1111/j.1083-6101.1997.tb00072.x

Longo, M. R., Schuur, F., Kammers, M. P., Tsakiris, M., & Haggard, P. (2008). What is embodiment? A psychometric approach. *Cognition, 107*(3), 978–998.

Lourenco, S. F., Longo, M. R., & Pathman, T. (2011). Near space and its relation to claustrophobic fear. *Cognition, 119*: 448–453.

Maes, P.-J., Leman, M., Palmer, C., & Wanderley, M. (2014). Action-based effects on music perception. *Frontiers in Psychology, 4* (1008). doi:10.3389/fpsyg.2013.01008

Maravita, A., & Iriki, A. (2004). Tools for the body (schema). *Trends in Cognitive Science, 8*(2), 79–86.

Maulucci, R. A., & Eckhouse, R. H. (2000). Retraining reaching in chronic stroke with real-time auditory feedback. *NeuroRehabilitation, 16*(3), 171–182.

McDonald, J. J., Teder-Salejarvi, W. A., & Hillyard, S. A. (2000). Involuntary orienting to sound improves visual perception. *Nature, 407*(6806), 906–908.

Menzer, F., Brooks, A., Halje, P., Faller, C., Vetterli, M., & Blanke, O. (2010). Feeling in control of your footsteps: Conscious gait monitoring and the auditory consequences of footsteps. *Cognitive Neuroscience, 1*(3), 184–192.

Merleau-Ponty, M. (1962). *Phenomenology of perception* (C. Smith, Trans.). London: Routledge & Kegan Paul.

Moseley, G. L., Gallace, A., & Spence, C. (2012). Bodily illusions in health and disease: physiological and clinical perspectives and the concept of a cortical "body matrix." *Neuroscience and Biobehavioral Reviews, 36*, 34–46.

Murray, C. D., Arnold, P., & Thornton, B. (2000). Presence accompanying induced hearing loss: Implications for immersive virtual environments. *Presence-Teleoperators and Virtual Environments, 9*(2), 137–148.

Neuhoff, J. G., Planisek, R., & Seifritz, E. (2009). Adaptive sex differences in auditory motion perception: Looming sounds are special. *Journal of Experimental Psychology: Human Perception and Performance, 35*, 225–234.

Newbold, J. W., Bianchi-Berthouze, N., Gold, N. E., Tajadura-Jiménez, A., & Williams, A. (2016). Musically informed sonification for chronic pain rehabilitation: Facilitating progress and avoiding over-doing. In *Proceedings of the 34th Annual ACM Conference on Human Factors in Computing Systems* (pp. 5698–5703). New York, NY: ACM.

Niedenthal, P. M. (2007). Embodying emotion. *Science, 316*, 1002–1005.

Niedenthal, P. M., Barsalou, L. W., Winkielman, P., Krauth-Gruber, S., & Ric, F. (2005). Embodiment in attitudes, social perception, and emotion. *Personality and Social Psychology Review, 9*(3), 184–211.

Nordahl, R. (2006). Increasing the motion of users in photo-realistic virtual environments by utilising auditory rendering of the environment and ego-motion. In M. Lombard & C. C. Bracken (Eds.), *Proceedings of the Ninth International Workshop on Presence* (pp. 57–63). International Society for Presence Research, Cleveland State University.

Olszewski, D. (2009). Targeted audio. In A. Waibel & R. Stiefelhagen (Eds.), *Computers in the human interaction loop* (pp. 133–141). London: Springer.

Pastore, R. E., Flint, J. D., Gaston, J. R., & Solomon, M. J. (2008). Auditory event perception: the source—perception loop for posture in human gait. *Perception and Psychophysics, 70*, 13–29.

Peters, E., Västfjäll, D., Slovic, P., Mertz, C. K., Mazzocco, K., & Dickert, S. (2006). Numeracy and decision making. *Psychological Science, 17*(5), 407–413.

Phillips, G. C., Jones, G. E., Rieger, E. J., & Snell, J. B. (1999). Effects of the presentation of false heart-rate feedback on the performance of two common heartbeat-detection tasks. *Psychophysiology, 36*(4), 504–510.

Pörschmann, C. (2000). Influences of bone-conduction and air conduction on the sound of one's own voice. *Acta Acustica, 86*, 1038–1045.

Rasamimanana, N., Bevilacqua, F., Schnell, N., Guedy, F., Flety, E., Maestracci, C.,… Petrevski, U. (2011). Modular musical objects towards embodied control of digital music. In *Proceedings of the Fifth International Conference on Tangible, Embedded, and Embodied Interaction*, *TEI '11* (pp. 9–12). doi:10.1145/1935701.1935704

Reidsma, D., Radha, M., & Nijholt, A. (2014). Mediated interactions and musical expression—A survey. In N. Lee (Ed.), *Digital Da Vinci* (pp. 79–98). New York, NY: Springer.

Rizzolatti, G. & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience 27*(1), 169–192.

Rizzolatti, G., Scandolara, C., Matelli, M., & Gentilucci, M. (1981). Afferent properties of periarcuate neurons in macaque monkeys. II. Visual responses. *Behavioural Brain Research, 2*, 147–163.

Roby-Brami, A., Van Zandt-Escobar, A., Jarrassé, N., Robertson, J., Schnell, N., Boyer, E.,… Bevilacqua, F. (2014). Toward the use of augmented auditory feedback for the rehabilitation of arm movements in stroke patients. *Annals of Physical and Rehabilitation Medicine, 57*, e4–e5.

Roginska, A. (2013). Auditory icons, earcons, and displays: Information and expression through sound. In S.-L. Tan, A. J. Cohen, S. D. Lipscomb, & R. A. Kendall (Eds.), *The psychology of music in multimedia* (pp. 339–364). Oxford: Oxford University Press.

Rosati, G., Rodà, A., Avanzini, F., & Masiero, S. (2013). On the role of auditory feedback in robotic-assisted movement training after stroke. *Computational Intelligence and Neuroscience, 11,* Article ID 586138.

Russo, C. R. (2015). The quantified self. In D. D. Schmorrow & C. M. Fidopiasti (Eds.), *Foundations of augmented cognition* (pp. 514–520). Cham: Springer International Publishing.

Schaffert, N., Mattes, K., & Effenberg, A. O. (2010). Listen to the boat motion: acoustic information for elite rowers. In *Proceedings of the Interactive Sonification (ISON) Workshop* (pages 31–38). Stockholm: KTH.

Schiphorst, T. (2011). Self-evidence: applying somatic connoisseurship to experience design. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems* (pp. 145–160). New York, NY: ACM.

Schneider, S., Münte, T. F., Rodriguez-Fornells, A., Sailer, M., & Altenmüller, E. (2010). Music supported training is more efficient than functional motor training for recovery of fine motor skills in stroke patients. *Music Perception: An Interdisciplinary Journal, 27*, 271–280.

Scholz, D. S., Wu, L., Pirzer, J., Schneider, J., Rollnik, J. D., Großbach, M., & Altenmüller, E. O. (2014). Sonification as a possible stroke rehabilitation strategy. *Frontiers in Neuroscience, 8*, 332.

Seth, A. K. (2013). Interoceptive inference, emotion, and the embodied self. *Trends in Cognitive Science, 17*(11), 565–573.

Sigrist, R., Rauter, G., Riener, R., & Wolf, P. (2013). Augmented visual, auditory, haptic, and multimodal feedback in motor learning: a review. *Psychonomic Bulletin and Review, 20*(1), 21–53.

Singh, A., Bianchi-Berthouze, N., & Williams, A. (2017). Supporting everyday function in chronic pain using wearable technology. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (pp. 3903–3915). doi:10.1145/3025453.3025947

Singh, A., Klapper, A., Jia, J., Fidalgo, A., Tajadura-Jiménez, A., Kanakam, N.,… Williams, A. (2014). *Motivating people with chronic pain to do physical activity: opportunities for technology design*. Paper presented at the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Toronto, Ontario, Canada.

Singh, A., Piana, S., Pollarolo, D., Volpe, G., Varni, G., Tajadura-Jiménez, A.,… Bianchi-Berthouze, N. (2015). Go-with-the-flow: Tracking, analysis and sonification of movement and breathing to build confidence in activity despite chronic pain. *Human-Computer Interaction 31*, 335–383.

Slater, M., & Usoh, M. (1994). Body centred interaction in immersive virtual environments. In N. M. Thalmann & D. Thalmann (Eds.), *Artificial Life and Virtual Reality* (pp. 125–148). New York, NY: John Wiley & Sons, Inc.

Sloboda, J. A. (2010). *Handbook of music and emotion: Theory, research, applications.* Oxford: Oxford University Press.

Smith, D. R. R., & Patterson, R. D. (2005). The interaction of glottal-pulse rate and vocal-tract length in judgements of speaker size, sex, and age. *Journal of the Acoustic Society of America, 118*(5), 3177–3186.

Sommer, R. (1959). Studies in personal space. *Sociometry, 22*, 247–260.

Soto-Faraco, S., & Väljamäe, A. (2012). Multisensory interactions during motion perception: From basic principles to media applications. In M. M. Murray & M. T. Wallace (Eds.), *The neural bases of multisensory processes* (Article ID: PMID:22593897). Boca Raton, FL: CRC.

Taffou, M., & Viaud-Delmon, I. (2014). Cynophobic fear adaptively extends peri-personal space. *Front Psychiatry, 5*, 122.

Tajadura-Jiménez, A., Basia, M., Deroy, O., Fairhurst, M., Marquardt, N., & Bianchi-Berthouze, N. (2015a). As light as your footsteps: Altering walking sounds to change perceived body weight, emotional state and gait. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 2943–2952). New York, NY: ACM.

Tajadura-Jiménez, A., Bianchi-Berthouze, N., Furfaro, E., & Bevilacqua, F. (2015b). Sonification of virtual and real surface tapping: Evaluation of behavior changes, surface perception and emotional indices. *IEEE MultiMedia, 22*, 48–57.

Tajadura- Jiménez, A., Kitagawa, N., Väljamäe, A., Zampini, M., Murray, M. M., & Spence, C. (2009). Auditory-somatosensory multisensory interactions are spatially modulated by stimulated body surface and acoustic spectra. *Neuropsychologia, 47*, 195–203.

Tajadura-Jiménez, A., Larsson, P., Väljamäe, A., Västfjäll, D., & Kleiner, M. (2010a). When room size matters: Acoustic influences on emotional responses to sounds. *Emotion, 10*(3), 416–422.

Tajadura-Jiménez, A., Liu, B., Berthouze, N., & Bevilacqua, F. (2014, October) *Altering touch behavior and perception with sonification*. Paper presented at the 8th Nordic Conference on Human-Computer Interaction, Helsinki, Finland.

Tajadura-Jiménez, A., Marquardt, T., Swapp, D., Kitagawa, N., & Bianchi-Berthouze, N. (2016). Action sounds modulate arm reaching movements. *Frontiers in Psychology, 7*, 1391.

Tajadura- Jiménez, A., Pantelidou, G., Rebacz, P., Västfjäll, D., & Tsakiris, M. (2011). I-space: the effects of emotional valence and source of music on interpersonal distance. *PLoS ONE, 6*(10), e26083.

Tajadura-Jiménez, A., Tsakiris, M., Marquardt, T., & Bianchi-Berthouze, N. (2015c). Action sounds update the mental representation of arm dimension: contributions of kinaesthesia and agency. *Frontiers in Psychology, 6*, 689.

Tajadura-Jiménez, A., Väljamäe, A., Asutay, E., & Västfjäll, D. (2010b). Embodied auditory perception: The emotional impact of approaching and receding sound sources. *Emotion, 10*(2), 216–229.

Tajadura-Jiménez, A., Väljamäe, A., Toshima, I., Kimura, T., Tsakiris, M., & Kitagawa, N. (2012). Action sounds recalibrate perceived tactile distance. *Current Biology, 22*(13), R516–517.

Tajadura-Jiménez, A., Väljamäe, A., & Västfjäll, D. (2008). Self-representation in mediated environments: the experience of emotions modulated by auditory-vibrotactile heartbeat. *Cyberpsychology and Behavior, 11*(1), 33–38.

Tonetto, P. L. M., Klanovicz, C. P., Spence, C. (2014). Modifying action sounds influences people's emotional responses and bodily sensations. *i-Perception*, 5153–5163.

Tsakiris, M. (2010). My body in the brain: a neurocognitive model of body-ownership. *Neuropsychologia, 48*(3), 703–712.

Tsakiris, M., & Haggard, P. (2005). Experimenting with the acting self. *Cognitive Neuropsychology, 22*(3), 387– 407.

Turchet, L., & Bresin, R. (2015) Effects of interactive sonification on emotionally expressive walking styles. *IEEE Transactions on Affective Computing, 6*, 152–164.

Turk, D. C., & Rudy, T. E. (1987). IASP taxonomy of chronic pain syndromes: preliminary assessment of reliability. *Pain, 30*(2), 177–189.

Väljamäe, A. (2009). Auditorily induced illusory self-motion: a review. *Brain Research Reviews, 61*(2), 240–255.

Väljamäe, A., & Kleiner, M. (2006). Spatial sound in auditory vision substitution systems. In *Proceedings of the 120th Audio Engineering Society Convention* (p. 6795). Paris: Audio Engineering Society.

Väljamäe, A., Larsson, P., Västfjäll, D., & Kleiner, M. (2008a). Sound representing self-motion in virtual environments enhances linear vection. *Presence, 17*(1), 43–56.

Väljamäe, A., Tajadura-Jiménez, A., Larsson, P., Västfjäll, D., & Kleiner, M. (2008b). Binaural bone-conducted sound in virtual environments: Evaluation of a portable, multimodal motion simulator prototype. *Acoustical science and Technology, 29*(2), 149–155.

van der Hoort, B., Guterstam, A., & Ehrsson, H. H. (2011). Being Barbie: The size of one's own body determines the perceived size of the world. *PLoS ONE, 6*(5), e20195.

Vázquez-Álvarez, Y., & Brewster, S.A. (2011). Eyes-free multitasking: the effect of cognitive load on mobile spatial audio interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2173–2176). New York, NY: ACM.

Viaud-Delmon, I., Mason, J., Haddad, K., Noisternig, M., Bevilacqua, F., & Warusfel, O. (2011). A sounding body in a sounding space: The building of space in choreography—focus on auditory-motor interactions [Supplemental material]. *Dance Research, 29*, 433–449.

Viaud-Delmon, I., & Warusfel, O. (2014). From ear to body: The auditory-motor loop in spatial cognition. *Frontiers in Neuroscience, 8*, 283.

Visell, Y., Fontana, F., Giordano, B. L., Nordahl, R., Serafin, S., & Bresin, R. (2009). Sound design and perception in walking interactions. *International Journal of Human-Computer Studies, 67*(11), 947–959.

Vogt, K., Pirro, D., & Kobenz, I. (2009). Physiosonic-movement sonification as auditory feedback. In *Proceedings of the 15th International Conference on Auditory Display, Copenhagen, Denmark* (pp. 1–7). Graz: IEM.

Wadley, G., Carter, M., & Gibbs, M. (2014). Voice in virtual worlds: The design, use and influence of voice chat in online play. *Human–Computer Interaction, 30*(3–4), 336–365.

Wallis, R. I., Ingalls, T., Rikakis, T., Olsen, L., Chen, Y., Xu, W., & Sundaram, H. (2007). Real-time sonification movement for an immersive stroke rehabilitation environment. In *Proceedings of the 13th International Conference on Auditory Display* (pp. 497–503). Montréal, Canada: Georgia Institute of Technology.

Wang, Q., Turaga, P., Coleman, G., & Ingalls, T. (2014, April). Somatech: an exploratory interface for altering movement habits. *In CHI'14 Extended Abstracts on Human Factors in Computing Systems* (pp. 1765–1770). New York: ACM.

Whalen, D. H., Hoequist, C. E., & Sheffert, S. M. (1995). The effects of breathe sounds on the perception of synthetic speech. *Journal of the Acoustical Society of America, 97*(5), 3147–3153.

Woll, S. B., & Mcfall, M. E. (1979). Effects of false feedback on attributed arousal and rated attractiveness in female subjects. *Journal of Personality, 4*(2), 214–229.

Wolpert, D. M., & Ghahramani, Z. (2000). Computational principles of movement neuroscience. *Neuroscience, 3*, 1212–1217.

Wolpert, D. M., Ghahramani, Z., & Jordan, M. I. (1995). An internal model for sensorimotor integration. *Science, 269*(5232), 1880–1882.

Zatorre, R. J., Chen, J. L., & Penhune, V. A. (2007). When the brain plays music: auditory-motor interactions in music perception and production. *Nature Reviews Neuroscience, 8*: 547–558.

Zotkin, D. N., Duraiswami, R., & Davis, L. S. (2004). Rendering localized spatial audio in a virtual auditory space. *IEEE Transactions on Multimedia, 6*(4), 553–564.

19

Input Device—Motion Capture

Atsushi Nakazawa and Takaaki Shiratori

Introduction

Motion-capture systems obtain a time series of the position and orientation of the objects, and then estimate the postures of articulated objects such as a human or animal's body. Since making the motions of high-dimensional articulated objects is quite difficult and time consuming for human designers, motion-capture systems are popularly used in a variety of applications, such as the movie and game industries, where realistic human or animals motions are necessary to control virtual actors or game characters. In robotics research, motion-capture data is used to control humanoid robots. In biology and medical applications, a motion-capture system is popularly used to evaluate the effects of rehabilitations and analyze human behaviors. In these traditional application fields, the motions of subjects were measured in a laboratory environment especially designed for motion capturing. Subjects were usually asked to wear a special suit with markers attached. The resulting data was quite accurate but the motion-capture system was expensive. Post data processing was also necessary to clean the data; realtime analysis was therefore usually difficult.

In recent years, thanks to the advance of computer vision and pattern-recognition technologies as well as new image sensors (such as real-time range image sensors), vision-based motion-capture systems have become practical. Compared to the traditional motion capture systems, users do not need to attach devices, so the latter systems are easy to use and cost effective. As a result, vision-based motion-capture systems are popularly used in interactive applications including game interfaces. The potential of these systems is increasing due to the numerous studies solving the existing drawbacks and, therefore, expanding the application fields. The organization of this chapter is as follows. In the second section, the structure of the human body in regard to motion-capture data is described. In the third section, several motion-capture approaches are introduced in terms of process flow, advantages, and disadvantages. Motion-capture databases and methods of reusing motion-capture data are then introduced.

The Wiley Handbook of Human Computer Interaction, Volume 1, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

Structure of the Human Body and Motion-capture Data

Structure of the human body

When motion capture is used, it is necessary to define the structure of the target object as a so-called skeleton. As the skeleton becomes more complicated, namely, the number of joints increases, more effort is required for measurement and postcleaning. It is therefore necessary to choose an appropriately simple skeleton for the target object considering the application tasks, particularly for interactive applications. Anatomically, the body of an adult human consists of 206 bones. It is composed of 270 bones at birth, then decreases to 206 bones by adulthood after some bones are fused together (Barnosky, 2010, p. 129). In addition, some of the bones are aligned in parallel (e.g., lower arm and shin) or tightly connected (e.g., torso and head); therefore, the number of movable parts is much smaller than the number of bones. Thus, the number of joints in the skeleton used in motion capture is much smaller than that of an actual human body. An example skeleton representing whole body is shown in Figure 19.1.

The definition of a human skeleton depends on the application task and the configuration of the motion-capture system. For example, if a whole-body character animation is to be generated, a whole-body skeleton must be used, and tiny portions, such as fingers and faces, must be omitted or simplified. On the contrary, when it is desired to capture motions during hand manipulations, tiny markers

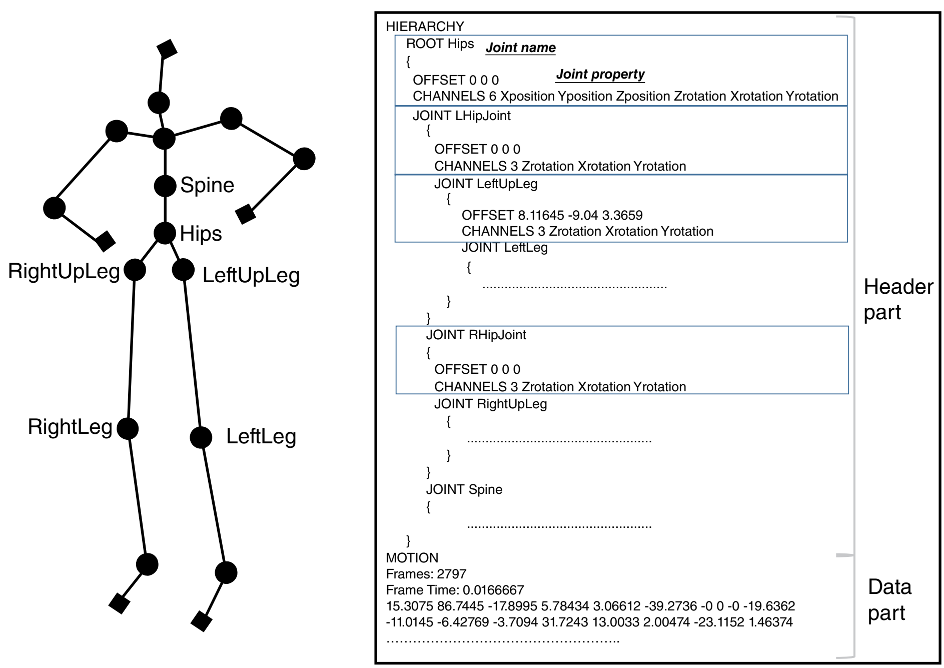


Figure 19.1 Typical whole-body model of motion capture data and corresponding BVH (biovision hierarchical data) file.

must be attached to each bone of the fingers, and the sensor configuration should be focused on the hands.

Motion-capture Data

A lot of file formats are used in motion-capture systems. Several major formats are listed in Table 19.1. Basically, they are categorized into two groups in terms of the kinds of data they hold. The first format type only holds 3D position data concerning joints or markers. Example file formats for this category include the C3D and TRC formats.2D position data are converted to 3D position data by using methods such as 3D reconstruction algorithms. Three-dimensional position data holds joint/marker positions and can therefore be used to render 3D graphics of static objects; however, it cannot be used for articulated character animation due to the lack of joint angles. It therefore requires a further joint angle calculation process to obtain joint angles from 3D marker points, and a skeleton model. The second format type holds the skeleton structure and time series of data that represent postures of the skeleton as well. These file formats include BVA/BVH, ASF/AMC and V/VSK formats. Thus, they can be used to animate the characters directly.

A standard whole-body skeleton model, which consists of 18 links and 14 joints, and the corresponding BVH (biovision hierarchical data) file are shown in Figure 19.1. The BVH format starts from a header part describing the skeleton, which describes the hierarchical structure of the articulated body with the joint names, properties of the joint (a list of parameters such as joint angles or positions), and the relative rotations and translations between the joints. After the header part, time-series of joint angles or position data follows. Here, one line of data corresponds to the one frame of captured data.

Table 19.1 Major file formats used in motion-capture systems.

Format name	*Description*	*Type*
BVA/BVH	Biovision hierarchical data. One of the standard formats representing humanlike characters motion.	Skeleton and joint angles
C3D	Developed by the National Institutes of Health in Bethesda, Maryland, open binary file format	Markers 3D position and associated data.
ASF/AMC	Developed by Acclaim. ASF (Acclaim Skeleton Format) contains a skeleton structure and AMC (Acclaim Motion Capture) holds movement data.	Skeleton and joint angles
TRC	Developed by Motion Analysis.	3D or 2D position
V/VSK	Vicon motion-capture file.	3D position, skeleton, joint angles
ASK/SDL	A variant of BVH file. ASK (Alias Skeleton) contains a skeleton structure and SDL contains movement data as well as many associated data.	Skeleton and joint angles

Methods of Motion Capture

As described in the introduction, a considerable number of motion-capture systems are commercially available. They have different features in terms of their accuracy, maximum frequency, usable situations, number of measurement points, freedom of marker arrangements, usability, and area of capturing fields. Major motion-capture approaches and their features are listed in Table 19.2.

Optical motion capture

Currently, optical motion capture is the most popular approach to obtain accurate and high-frequency human motions; therefore, many commercial systems are available. The idea behind an optical system is to obtain 3D positions of the markers attached to the users body by using triangulation technique from multiple camera images. The original idea of Rashid was to use light bulbs (Rashid, 1980); however, major systems use reflective markers and IR cameras fitted with IR lights around the camera lens (Figure 19.2a). The typical flow of the optical motion capture is explained as follows.

At first, it is necessary to perform *camera and system calibrations.* Usually, these steps are carried out by showing a single marker or several markers attached to a known-sized bar (wand) to multiple cameras. As a result, the relative positions of the cameras and a ground plane are obtained. Next, *skeleton capturing and calibration* is conducted. A subject is asked to wear a motion-capture suit with reflective markers attached (Figures 19.2b–d). As a marker of optical motion capture only produces 3D position information, it needs more than three markers for each bone to obtain 3D position and orientation information. After that, the subject moves the joints and the system obtains a range of motion and bone lengths. After the capturing session, *post processing (data cleaning)* needs to be performed. Because the initial measurement results (3D-marker positions) are erroneous due to the spatio matching error in the stereo rig and temporal matching error between close markers, so-called marker swapping, manual correction is conducted to remove the errors.

Advantages and disadvantages Optical motion-capture systems have a long history, so their methods are well studied and matured. Their accuracy is good (less than 1 cm), and they allow high-frequency measurement (more than 1000 Hz in several systems). Another important property is their applicability to a variety of targets—not only to the human body but also to faces, animals or soft / rigid objects, by changing the skeletons. In case of facial capture, tiny markers are attached to the facial keypoints, and these are used to deform the 3D facial models. Due to these performance merits and freedom of applications, optical motion capture is the most popular, and is used for a lot of applications, including movies, games, and medical and biokinematics fields.

On the other hand, optical motion capture has some restrictions when it is used for user interfaces and realtime applications due to its poor usability (the need for calibrations and wearing motion-capture suits) and marker-swapping errors. It also requires a larger system space than the capturing area, so the markers are observed by the multiple cameras. In addition, the system cost is still high (though low-cost systems have become available in recent years).

Table 19.2 Motion-capture systems and their features.

Type of system	*Accuracy*	*Max. frequency*	*System size*	*Number of measurement points*	*Freedom of arrangement*	*Usability*	*Supported field*
Optical	High	High	Large	Large	High	Low	Limited
Magnetic	High	Medium	Large	Medium	Medium	Low	Limited
Vision	Low	Medium	Small	Low-medium	Low-medium	High	Limited, unlimited
Mechanical	Variable	High	Small	Low	Low	Low	Unlimited
Wearable sensors (IMU, body-attached vision)	Low	Medium	Small	Low	Variable	Medium	Unlimited
Instrumented puppet	Variable	High	Small	Low	Low	High	Limited, unlimited

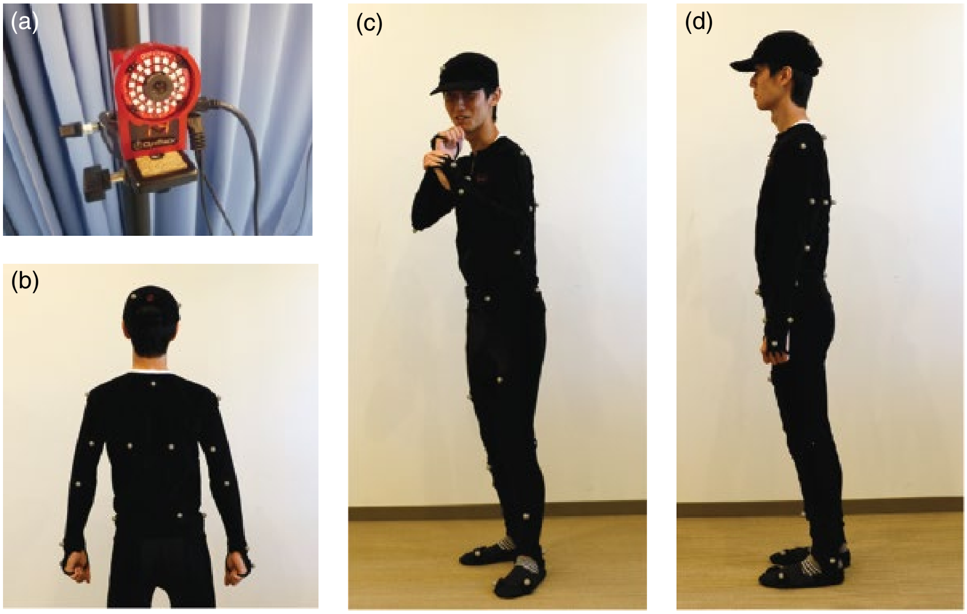


Figure 19.2 A typical optical motion-capture system. (a) An infrared (IR) camera with IR LEDs; (b)–(d) a motion capture suit where reflective markers are attached.

Magnetic motion capture

Magnetic motion capture is another major approach and is used for many applications like the optical approach. The idea behind magnetic motion capture is to use an artificially generated magnetic field, and magnetic position and direction sensors attached to targets such as human bones. The typical flow of magnetic motion capture is as follows. In a similar manner to the optical approach, *calibration* is necessary before the capturing sessions. That is, magnetic fields in the capturing area are measured by placing a calibration pole, to which several magnetic sensors are attached, at grid points. Then, *skeleton capturing* is performed. In the capturing session, an actor needs to wear a motion-capture suit with magnetic sensors attached. Since a marker of magnetic motion capture outputs position and orientation information at the same time, the number of required markers is small in contrast to the optical motion capture. The sensor readings are transmitted to the system, and the pose of the skeleton is reconstructed. The effort required for postprocessing (data cleaning) is smaller than that required by optical systems as it does not have errors in stereo reconstruction and temporal correspondences.

Advantages and disadvantages Like optical systems, magnetic motion-capture systems are used in many applications fields. Their advantages compared to optical systems are threefold: postprocessing is not required or is very easy, the system is smaller, and position and orientation value can be obtained simultaneously. It is therefore suitable for realtime applications such as avatar control for live broadcasts. However, it still requires calibrations and an actor wearing a capture suit. The number of measurement points is usually smaller than that for optical systems but the cost is high.

Markerless vision motion capture

The idea behind vision-based motion capture is to use only a vision sensor (camera) for capturing a human posture. This type of capture uses single or multiple cameras to take a video or a single frame of a subject and reconstruct 3D postures.

In computer-vision research fields, this technique is so-called figure tracking, on which a considerable number of works has been reported (Khan & Shah, 2009; Moeslund & Granum, 2001). In the early 1990s, Hogg proposed a deformable 3D stick-figure model whose parameters are estimated so that it fits to the input images (Hogg, 1983). However, because 3D posture reconstruction from a single 2D image is fundamentally an ill-posed problem, several assumptions, such as supervised information (manual annotations) at keyframes or introducing human-motion priors (Brubaker, Fleet, & Hertzmann, 2007; Urtasun, Fleet, & Fua, 2006), are necessary to obtain the parameters.

To solve the issue described above, other studies have uses multiple images or range images for 3D posture estimation. Several works expanded Hoggs stick-figure fitting using multiple images to solve the ambiguity issue, and other efforts first reconstruct 3D volume data and estimate 3D postures. The most successful approach of recent years uses realtime depth (range) image sensors. The Microsoft Kinect sensor (Zhang, 2012) outputs realtime range images using a projected random dot pattern or time-of-flight information. From the output, 3D posture is estimated by methods such as random forest, which can output the parameters much faster than model-fitting approaches. Because commodity depth sensors cannot be used under sunlight, which contains a significant amount of infrared light, RGB cameras are preferred to capture outdoor motion. Recent studies represent human body shape as many 3D Gaussian spheres along a skeleton for outdoor markerless motion capture (Elhayek, Stoll, Kim, & Theobalt, 2015). This representation enables an analytically differentiable objective function and therefore robust estimation of human skeletal poses with only a few cameras.

Advantages and disadvantages The key advantages of vision-based systems are superior usability and low system cost. Recent vision-based systems do not use wearable suits or other equipment. The sensor parameters are usually calibrated when it is installed; therefore, these systems usually do not need any preparations. In addition, it only uses cameras or range-image sensors available on the market, so the system cost is becoming smaller and smaller. For these reasons, vision-based systems are suitable for user interfaces and, in reality, are popularly used in consumer and experimental devices.

However, the accuracy of the estimation by vision-based systems is limited compared to the optical and magnetic systems (which use markers). For example, vision-based systems cannot easily estimate rotation of the arms or the cases that some portions are occluded or merged together; vision systems are therefore rarely used for applications in which professional contents are created.

Other approaches

The three approaches described above assume the sensors, such as cameras or magnetic-field generators, are fixed to an environment; thus, they cannot be used for mobile motion capture. For such situations, the following approaches have been developed.

Mechanical motion capture In mechanical motion capture, mechanical sensors are attached to an actor's body, and its 3D postures are estimated from the sensor reads. The sensors used include angular sensors (potentiometers), gyros, and shape sensors. An actor needs to wear these sensors; therefore, the usability of mechanical motion capture is limited, but it can support mobile measurement. This advantage is important in regard to expanding the application field of motion capture to, for example, sports and factory scenes.

IMU-based motion captures Motion captures based on inertial measurement units (IMUs) use body-attached IMUs to obtain a body configuration. The IMU consists of accelerometers and gyroscopes, and estimates the markers' position and orientation by accumulating time series of velocity, orientation, and gravitational forces. Similar to the mechanical motion capture, IMU supports mobile measurement but sensor size is much smaller. Thus these have increased usability (Figure 19.3). However, as it integrates time-series of measurement, it potentially has a drift issue. Several latest efforts solve this issue by combining other type of sensors (Vlasic et al., 2007).

Body-attached vision motion captures This type of motion-capture system uses body-attached vision sensors. The Prakash system uses specially designed body-attached tags that can read time-coded pattern light by using a single photo sensor. A subject is illuminated by the coded pattern light projected by a specially designed projector. During measurement, the body-attached badges read the coded pattern and obtain 3D position with respect to the projector. As a result, all 3D positions of the badges are obtained, and the subject's posture is constructed. In demonstration, mobile motion capture was achieved by using a projector to illuminate a running subject.

Figure 19.3 Motion-capture system using wearable inertia sensors (Xsens systems).

A more recent system uses small recordable cameras as the body-attached sensors (Shiratori, Park, Sigal, Sheikh, & Hodgins, 2011) (Figure 19.4). From each image, 3D positions and orientations are obtained by using a structure-from-motion-like approach. Namely, it detects natural feature points of the surrounding environment from the input video and tracks them. Accordingly, 3D positions and orientations of the camera are obtained. Combining estimation results from all cameras as well as the kinematic constraints between cameras in consideration of human bone structure, a human posture is reconstructed. Although it requires significant computational cost, and the cameras are larger than inertia sensors, it supports mobile capturing and obtains global positions of the subject from the 3D geometry of the surrounding environment.

Instrumented puppet Motion-capture data can be generated not only by taking human motion directly but also by using humanlike figures (instrumented puppet interfaces) whose joint angles or positions are measured by a variety of sensors similar

Figure 19.4 Motion-capture system using body-attached cameras. It reconstructs the subject's posture and scene geometry simultaneously.

to other motion-capture systems (such as vision and mechanical sensors). Esposito created an instrumented puppet called the Monkey that allowed users to specify human or character poses intuitively (Esposito, Paley, & Ong, 1995). A similar instrumented puppet was developed by Knep to specify poses of dinosaurs for movie production (Knep, Hayes, Sayre, & Williams, 1995). Johnson developed an instrumented puppet to control a birdlike character (Johnson, Wilson, Blumberg, Kline, & Bobick, 1999). The user's manipulation of the puppet was recognized by using hidden Markov models (HMMs) to select from a set of predetermined motion patterns. In more recent years, thanks to the technology of smaller humanoid robots, smarter systems have been developed (Figure 19.5) (Yoshizaki et al., 2011; Numaguchi, Nakazawa, Shiratori, & Hodgins, 2011), and some of them, such as that reported by CELSYS (2016), which is designed for use in computer animations, are commercially available.

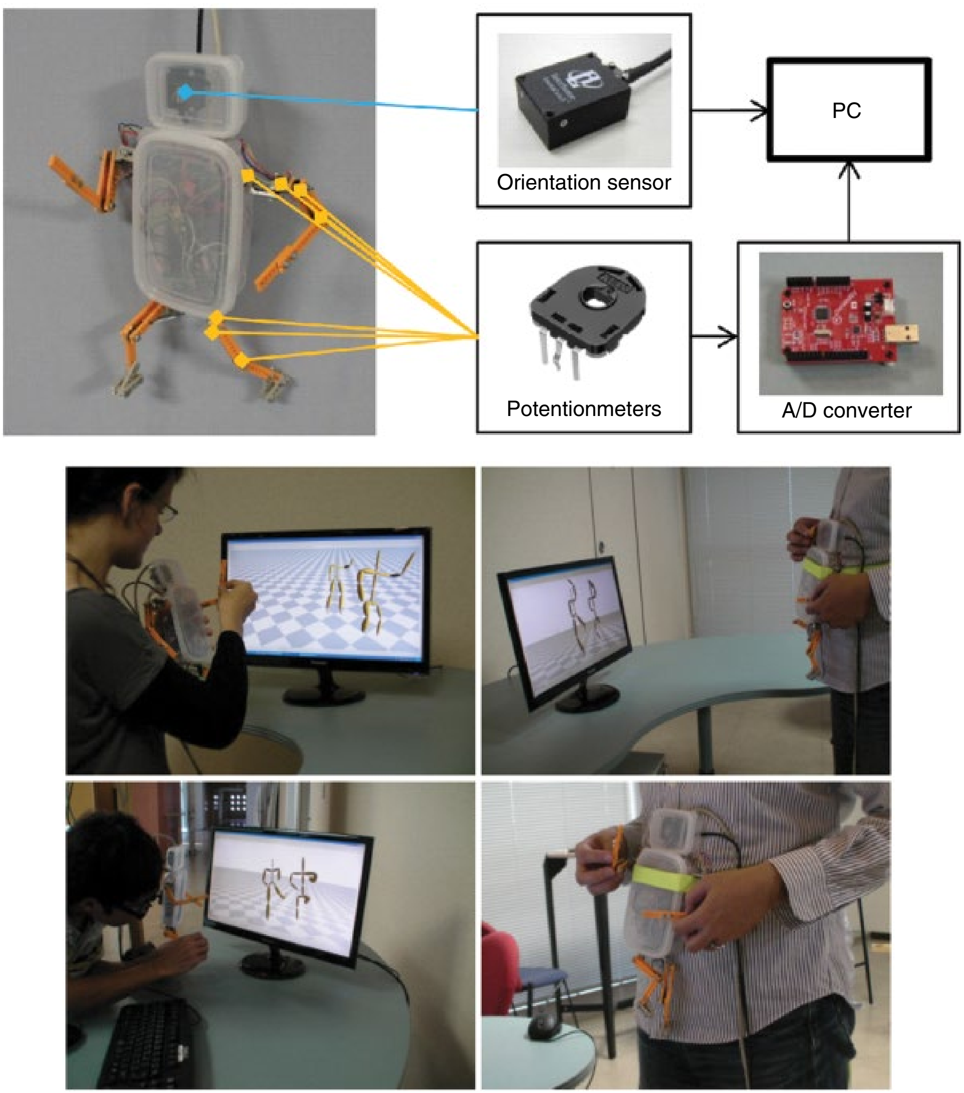


Figure 19.5 Puppet interface for motion capture retrieval.

Because these toy puppets do not capture human-body movement directly, the reality of the captured data is limited, even though the puppets are easy to use and portable. They can also generate motions that are difficult for human actors to perform (such as backflips or falling down stairs).

Reuse of Motion-capture Data: Motion-capture Databases and their Applications

Although many kinds of motion-capture devices have become commercially available, they still require a considerable amount of effort and cost to capture and clean the captured data. Thus, reuse of precaptured data has been gaining interest in the motion-capture industry, and several studies on generating new motions from precaptured data have been conducted. The relations between motion-capture databases and their applications are shown in Figure 19.6.

Motion-capture Databases

Major open motion-capture databases currently maintained are listed in Table 19.3. Several of them have search function using keywords (e.g., CMU motion-capture database) or video annotations. The file type depends on the database but several extensive communities have converted to other formats such as BVH.

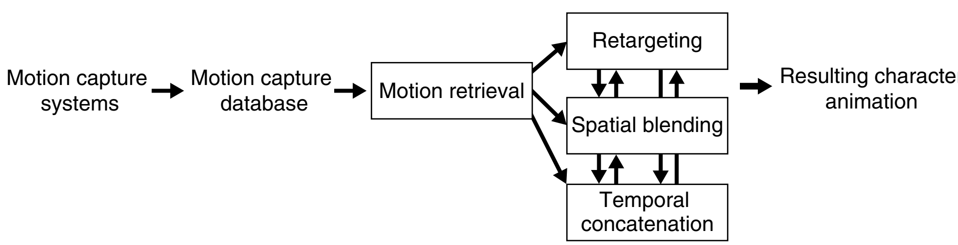


Figure 19.6 Motion-capture database and its application techniques.

Table 19.3 List of major motion-capture databases.

Name	*Size (as of Sep. 2015)*	*File type*	*URL*
CMU Graphics Lab motion-capture database	2605 trials, 23 subjects, 6 categories	tvd, c3d, amc mpg	http://mocap.cs.cmu.edu/
Motion- Capture Database HDM05	5 subjects, 5 categories	c3d, asfamc	http://resources.mpi-inf.mpg.de/HDM05/
HumanEva Dataset	6 subjects, 6 common actions	c3d, video from multiple viewpoints	http://humaneva.is.tue.mpg.de/
Mocapdata.com	More than 5000	bvh, c3d	http://www.mocapdata.com/

Retrieval of motion-capture data

To maintain and reuse motion-capture data, retrieval of motion capture data is an essential technique. Without such an efficient search technique, the database is difficult to maintain and unused. Although most motion-capture databases use verbal annotations (keywords), keywords cannot describe detailed features of the motion. For example, if the database includes many kinds of dance motions, it is difficult to find the desired one using only keywords.

Developing efficient motion retrieval depends on finding a similarity metric of human postures, but that is quite a difficult task. In the computer graphics community, several efforts have been made to solve this issue. Bruderlin and Williams (1995) demonstrated that dynamic time warping (DTW) can be used to evaluate the similarity between two motion-capture sequences. Müller, Röder, and Clausen (2005) developed a content-based method for motion retrieval. With their method, human motion is abstracted with geometrical relationships of body-part pairs. Ishigaki built a performance interface that translated the user's intentions to human motion by applying a subspace method with principal component analysis (PCA) to a motion-capture database (Ishigaki, White, Zordan, & Liu, 2009). Extending the PCA-based approach, Numaguchi et al. (2011) proposed using a mutual-subspace method (MSM) and a dual-subspace projection method (DSPM) combined with their puppet interface and build an efficient motion-search algorithm. Ho and Komura (2009) retrieved interactions between characters on the basis of topological information derived from knot theory.

Retargeting

Retargeting is a technique that applies precaptured motion to other (target) characters while preserving the feature of the original motion (see Figure 19.7). When motion-capture data is applied to a target character whose body size is different from that of the motion-capture data, the resulting character animation becomes unnatural, as in the "foot-skating" effect (i.e., the feet do not touch the ground).

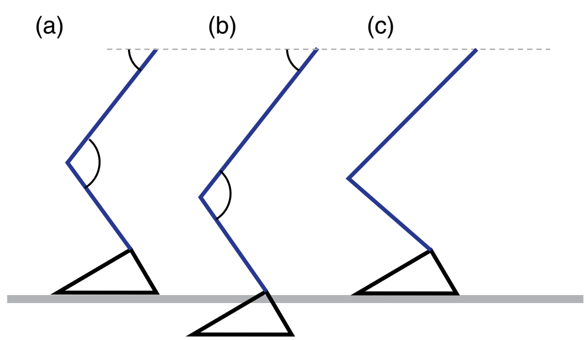


Figure 19.7 An example of motion retargeting. (a) Original motion capture data; (b) foot penetration happens when the joint angles of (a) are directly transferred to the skeleton with different link lengths; (c) the motion retargeting technique solve the problem modifying the original posture to the target skeleton so that it satisfies the foot contacting state as the same as (a).

To solve this issue, Gleicher proposed a general framework for motion retargeting. In their framework, first they identify the constraints on the original motion (such as the feets contact points with the ground or the positions of clapping hands) and apply them to the target character. Then, the target motion is modified so that it satisfies the constraints of the original motion. The optimization method used is similar to the space-time constraint (Lee & Shin, 1999; Shin, Lee, Shin, & Gleicher, 2001; Witkin & Kass, 1988).

Motion editing

As well as motion retargeting, motion editing is an essential technique for producing the desired motion. As illustrated in Figure 19.6, motion editing can be categorized as two types of techniques, namely, spatial blending, which produces new motion by blending several motions on the same timeline, and temporal concatenation, which generates temporally different motion or longer sequences by splitting and concatenating motion-capture data.

Spatial blending is used for slight modification from the original motions while preserving the features of original behavior, such as avoiding obstacles while walking or reaching actions to arbitrary targets. It includes techniques using signal processing (Bruderlin & Williams, 1995), warping the motion to satisfy constraints, such as a particular joint reaching the desired position at the desired time (Brand & Hertzmann, 2000), learning motion patterns for extracting style components (Grochow, Martin, Hertzmann, & Popović, 2004; Hsu, Pulli, & Popović, 2005; Witkin & Popovic, 1995), posture constraints (Yamane, Kuffner, & Hodgins, 2004), or inverse kinematics using several example motions to generate the desired hand motion while preserving motion styles (Bruderlin & Williams, 1995).

Although temporal concatenation also uses a motion-capture database, it generates new motions by splitting and concatenating several motions. It can thus generate longer sequence than the original data. One of the representative methods is Motion Graph (Kovar, Gleicher, & Pighin, 2002). First, this finds similar frames (postures) among many motion-capture data. Then, it connects the frames and constructs a graph structure. As a result, new motion is generated by tracing the graph. Considering constraints for graph traversal, desired motions, such as the users' input, generated paths, environmental constraints and other signals, such as musical beats, can be synthesized (Arikan & Forsyth, 2002; Lee, Chai, Reitsma, Hodgins, & Pollard, 2002; Li, Wang, & Shum, 2002; Pullen & Bregler, 2002; Shiratori, Nakazawa, & Ikeuchi, 2006).

Summary

In this chapter, approaches to motion capture were briefly introduced. Many motion-capture approaches have been proposed but the appropriate one should be chosen in consideration of the constraints of the application, such as accuracy, requirements for realtime and mobile capturing, and system size and cost. After that, techniques for treating motion-capture data, including methods for retrieving motion data as well as retargeting and editing algorithms, were summarized.

References

Arikan, O., & Forsyth, D. A. (2002). Interactive motion generation from examples. *ACM Transactions on Graphics, 21*, 483–490.

Barnosky, A. (2010). *Mammal anatomy: An illustrated guide*. New York, NY: Marshall Cavendish.

Brand, M., & Hertzmann, A. (2000). Style machines. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques* (pp. 183–192). doi:10.1145/344779.344865

Brubaker, M. A., Fleet, D. J., & Hertzmann, A. (2007). Physics-based person tracking using simplified lower-body dynamics. In *Proceedings of the 2007 IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1–8).

Bruderlin, A., & Williams, L. (1995). Motion signal processing. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques* (pp. 97–104). doi:10.1109/CVPR.2007.383342

CELSYS. (2016). *Qumarion*. Retrieved from http://www.clip-studio.com/quma/en/

Elhayek, A., Stoll, C., Kim, K. I., & Theobalt, C. (2015). Outdoor human motion capture by simultaneous optimization of pose and camera parameters. *Computer Graphics Forum, 34* (6), 86–98.

Esposito, C., Paley, W. B., & Ong, J. (1995). Of mice and monkeys: A specialized input device for virtual body animation. In *Proceedings of the 1995 Symposium on Interactive 3D Graphics* (pp. 109–114). doi:10.1145/199404.199424

Grochow, K., Martin, S. L., Hertzmann, A., & Popović, Z. (2004). Style-based inverse kinematics. *ACM Transactions on Graphics, 23*, 522–531.

Ho, E. S., & Komura, T. (2009). Indexing and retrieving motions of characters in close contact. *IEEE Transactions on Visualization and Computer Graphics, 15*(3), 481–492.

Hogg, D. (1983). Model-based vision: A program to see a walking person. *Image and Vision Computing, 1*(1), 5–20.

Hsu, E., Pulli, K., & Popović, J. (2005). Style translation for human motion. In *ACM Transactions on Graphics, 24*, 1082–1089.

Ishigaki, S., White, T., Zordan, V. B., & Liu, C. K. (2009). Performance-based control interface for character animation. *ACM Transactions on Graphics, 28*, 61.

Johnson, M. P., Wilson, A., Blumberg, B., Kline, C., & Bobick, A. (1999). Sympathetic interfaces: Using a plush toy to direct synthetic characters. In *Proceedings of the SigChi Conference on Human Factors in Computing Systems* (pp. 152–158). doi:10.1145/302979.303028

Khan, S. M., & Shah, M. (2009). Tracking multiple occluding people by localizing on multiple scene planes. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 31*(3), 505–519.

Knep, B., Hayes, C., Sayre, R., & Williams, T. (1995). Dinosaur input device. In *Proceedings of the SigChi Conference on Human Factors in Computing Systems* (pp. 304–309). doi:10.1145/223904.223943

Kovar, L., Gleicher, M., & Pighin, F. (2002). Motion graphs. In *ACM Transactions on Graphics, 21*, 473–482.

Lee, J., Chai, J., Reitsma, P. S., Hodgins, J. K., & Pollard, N. S. (2002). Interactive control of avatars animated with human motion data. In *ACM Transactions on Graphics, 21*, 491–500.

Lee, J., & Shin, S. Y. (1999). A hierarchical approach to interactive motion editing for human-like figures. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques* (pp. 39–48). doi:10.1145/566570.566607

Li, Y., Wang, T., & Shum, H.-Y. (2002). Motion texture: a two-level statistical model for character motion synthesis. *ACM Transactions on Graphics, 21*, 465–472.

Moeslund, T. B., & Granum, E. (2001). A survey of computer vision-based human motion capture. *Computer Vision and Image Understanding, 81*(3), 231–268.

Müller, M., Röder, T., & Clausen, M. (2005). Efficient content-based retrieval of motion capture data. *ACM Transactions on Graphics, 24*, 677–685.

Numaguchi, N., Nakazawa, A., Shiratori, T., & Hodgins, J. K. (2011). A puppet interface for retrieval of motion capture data. In *Proceedings of the 2011 ACM Siggraph/Eurographics Symposium on Computer Animation* (pp. 157–166). New York, NY: ACM.

Pullen, K., & Bregler, C. (2002). Motion capture assisted animation: Texturing and synthesis. In *ACM Transactions on Graphics, 21*, 501–508.

Rashid, R. F. (1980). Towards a system for the interpretation of moving light displays. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 6*, 574–581.

Shin, H. J., Lee, J., Shin, S. Y., & Gleicher, M. (2001). Computer puppetry: An importance-based approach. *ACM Transactions on Graphics, 20*(2), 67–94.

Shiratori, T., Nakazawa, A., & Ikeuchi, K. (2006). Dancing-to-music character animation. *Computer Graphics Forum, 25*, 449–458.

Shiratori, T., Park, H. S., Sigal, L., Sheikh, Y., & Hodgins, J. K. (2011). Motion capture from body-mounted cameras. *ACM Transactions on Graphics, 30*, 31.

Urtasun, R., Fleet, D. J., & Fua, P. (2006). 3D people tracking with Gaussian process dynamical models. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)* (Vol. *1*, pp. 238–245). doi:10.1109/CVPR.2006.15

Vlasic, D., Adelsberger, R., Vannucci, G., Barnwell, J., Gross, M., Matusik, W., & Popović, J. (2007). Practical motion capture in everyday surroundings. In *ACM Transactions on Graphics, 26*, 35.

Witkin, A., & Kass, M. (1988). Spacetime constraints. *ACM Siggraph Computer Graphics, 22*(4), 159–168.

Witkin, A., & Popovic, Z. (1995). Motion warping. In *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques* (pp. 105–108). doi:10.1145/218380.218422

Yamane, K., Kuffner, J. J., & Hodgins, J. K. (2004). Synthesizing animations of human manipulation tasks. In *ACM Transactions on Graphics, 23*, 532–539.

Yoshizaki, W., Sugiura, Y., Chiou, A. C., Hashimoto, S., Inami, M., Igarashi, T.,… Mochimaru, M. (2011). An actuated physical puppet as an input device for controlling a digital manikin. In *Proceedings of the SigChi Conference on Human Factors in Computing Systems* (pp. 637–646). doi:10.1145/1978942.1979034

Zhang, Z. (2012). Microsoft Kinect sensor and its effect. *IEEE multimedia, 19*(2), 4–10.

20

Applications of Intelligent and Multimodal Eye-Gaze Controlled Interfaces

Pradipta Biswas and Pat Langdon

Introduction

Recent advances in infrared-based eye-gaze trackers have significantly increased research and industrial use of gaze-tracking technology. Although research involving analysis of eye-gaze dates back to the early 19th century, until recently, eye-gaze trackers were mainly used for analyzing ocular parameters for reading and a variety of human-machine interaction tasks. With progress in processor speed and image-processing algorithms, it is now also possible to use gaze-tracking technology in real time to control a screen pointer in a direct-manipulation interface. Gaze-controlled interfaces have already been investigated and used for assistive technology, and in automotive and aviation environments. This chapter presents a set of systems to improve quality of interaction in eye-gaze controlled interfaces.

In a two-dimensional screen, we mainly investigate saccadic and small-pursuit eye-gaze movements. Saccadic movement takes 250 to 350 ms to complete and is ballistic in nature. However, small-pursuit movements keep the eye-gaze moving around the point of interest. If we move a pointer directly following eye-gaze, the small-pursuit movement creates jitter and it becomes difficult to select a target if the pointer is not stable.

The best available accuracy of eye-gaze tracker is 0.4° of the visual angle as of February, 2015. This accuracy translates to approximately 18 pixels to a standard desktop screen from 65 cm of viewing distance. So a gaze-control interface may occasionally need a user to focus a little bit off target to bring the cursor onto a screen element.

Overall, it is not difficult to move a screen pointer based on eye-gaze but focusing the screen pointer on a screen element remains a challenge in a gaze-controlled interface. Existing gaze-controlled software solves this issue by designing special interfaces with big screen elements to compensate for variation and limitations in accuracy. However, interaction systems should not limit interface design and should work for existing interfaces without limiting the size of screen elements.

The Wiley Handbook of Human Computer Interaction, Volume 1, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

Our research tried to reduce pointing and selection times as well as cognitive load in gaze-controlled interfaces in two ways. We have developed a target prediction and expansion technology that can activate a target without needing the pointer to reach on top of the target. Secondly, we have combined other input modalities with gaze-tracking to help in pointing and selection. This chapter presents the following case studies of using intelligent and multimodal gaze-tracking technology:

- a pointing and selection task in a graphical user interface involving an eye-gaze tracker, joystick, and Leap motion controller;
- browsing Google Maps using eye-gaze tracking;
- shopping electronically using intelligent eye-gaze tracking by novice computer users;
- controlling a dashboard (stack) in a driving simulator;
- controlling a multifunction display in the cockpit of a flight simulator.

New Gaze-Tracking Technologies

Researchers have already investigated combining gaze-tracking with other input modalities. The MAGIC pointing system (Zhai, Morimoto, & Ihde, 1999) explored combining the use of the mouse with eye-gaze tracking-based pointing. The recent Tobii EyeX system also provides a similar feature for using eye-gaze with touchpad or mouse-based pointing. Bates (1999) combined a Polhemus tracker with eye-gaze tracking-based pointing and their multimodal eye-tracking system allows for the zooming of a portion of a screen using a Polhemus tracker. Zandera, Gaertnera, Kothea, and Vilmek (2010) combined a BCI system with eye-gaze tracking, where EEG generated by imagining a rinsing action is trained to make a selection. However, their system had limited success in reducing pointing times.

Eye-gaze tracking with scanning

This system was developed to integrate eye-gaze tracking with assistive technology to help in pointing and selection tasks by people with motor impairment.

Initially, the system moves the pointer across the screen based on the eye-gaze of the user (Figure 20.1). Users can see a small button moving across the screen and the button is placed approximately where they are looking at the screen. We extract the eye-gaze position by using the SDK provided with an eye-gaze tracker and used a median filter that changes the pointer position every 500 ms. The users can switch to the scanning system by blinking or giving a key press anytime during eye-tracking. The duration of blink is configurable to distinguish between intentional and unconscious blinks.

We have used a particular type of scanning system, known as eight-directional scanning to navigate across the screen. In the eight-directional scanning technique, the pointer icon is changed at regular time intervals to show one of eight directions (up, up left, left, left down, down, down right, right, right up). The user can choose a direction by pressing a switch or blinking when the pointer icon shows the required

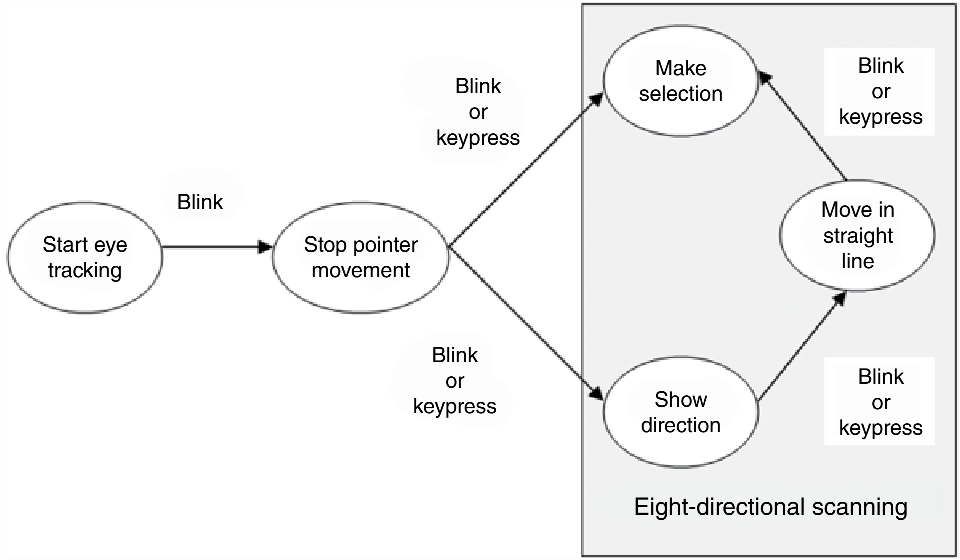


Figure 20.1 State chart diagram of eyegaze-scanning combined system.

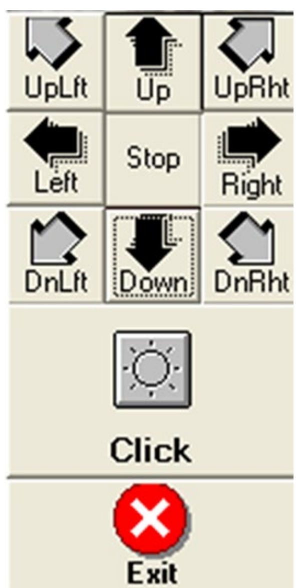


Figure 20.2 Screenshot of scanning system.

direction. After getting the direction choice, the pointer starts moving. When the pointer reaches the desired point on the screen, the user has to make another key press to stop the pointer movement and make a click. The user can move back to the eye-gaze tracking system from the scanning system by selecting the exit button in the scanning interface (Figure 20.2).

A demonstration of the scanning system can be seen at http://www.youtube.com/watch?v=0eSyyXeBoXQandfeature=user. A couple of videos of the system can be found from the following links:

- screenshot: http://www.youtube.com/watch?v=UnYVO1Ag17U
- actual usage: http://www.youtube.com/watch?v=2izAZNvj9L0

Our study (Biswas & Langdon, 2011) confirmed that the technique is faster than only a scanning-based interface as users can move the pointer through a large distance on the screen using their eye-gaze quicker than using only a single-switch scanning interface.

Eye-gaze tracking with joystick

This system is developed to integrate eye-gaze tracking with a joystick mainly for the military aviation environment. The pointer is initially moved in screen based on the user's eye-gaze, however if the user moves the joystick, the eye-gaze tracking based pointing is switched off and the pointer moves are based on joystick input. The joystick button is used for selection, and once a selection is made the eye-gaze tracking based pointing turns on again (Figure 20.3).

Eye-gaze tracking with Leap motion

In this technique, we used a Leap-motion controller with the eye-gaze tracker. The Leap motion controller is used to make small corrective movements when the eye-gaze tracker alone could not bring the pointer on target. If the user puts his hand on the Leap-motion sensor, the pointer stops moving based on eye-gaze. We used the logarithm of change in finger position to move the pointer based on hand movement. The logarithm function ensured the pointer did not move more than 1° of the visual angle from the previous position so that hand movement could only be used for the homing phase, and not used for ballistic movement. When the user removed his hand from top of the Leap-motion sensor, the pointer resumed to move based on the eye-gaze of the user.

We used the left mouse button for selection, although the light sensor in the mouse was blocked to ensure the screen pointer did not move following mouse movement. A demonstration video of the system in the context of Web browsing can be found at http://youtu.be/AnAZxJ6U9Wc. See Figure 20.4.

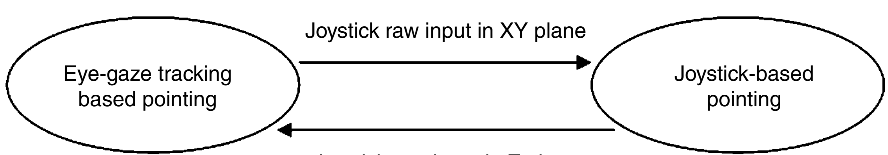


Figure 20.3 State chart diagram of eyegaze-joystick combined system.

Intelligent gaze tracking—target prediction

Unlike in the previous three cases, we did not combine any other pointing modality for this particular technology; rather, we tried to predict and expand users' intended targets. In a two-dimensional screen, while people search for an item, they usually make a big saccadic movement towards the target and then a set of smooth pursuit movements to visually investigate the target. These two phases of movements roughly correspond to the ballistic and homing phases of a rapid aiming movement.

We developed a neural network-based model (Figure 20.5) that takes different trajectory profiles, like velocity, acceleration, and bearing of movement as input parameters and, based on that, predicts the type of eye movement. If the model predicts small amplitude jittery movements, we assume the user is already near to his intended target. Then we expand the nearest target to one-and-a-half times its original size from the user's present gaze location.

Researchers already explored similar technology in terms of intent recognition or next point prediction (Lank, Cheng, & Ruiz, 2007; Ziebart, Dey, & Bagnell, 2012), but we pioneered the use of similar technology for gaze-controlled interfaces. Our user studies (Biswas & Langdon, 2015) confirmed that the model can significantly reduce pointing and selection times in gaze-controlled interfaces. A demonstration video of the target prediction technology can be found at https://www.youtube.com/watch?v=p9YOKj59TiY.

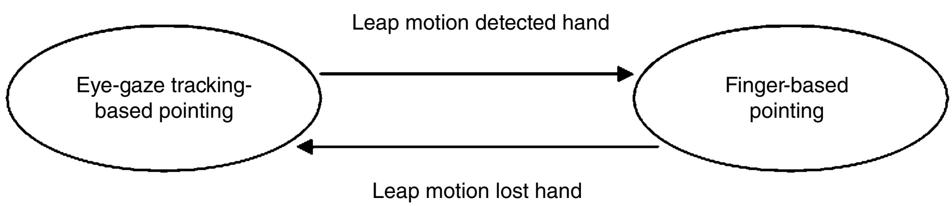


Figure 20.4 State chart diagram of eyegaze-leap motion combined system.

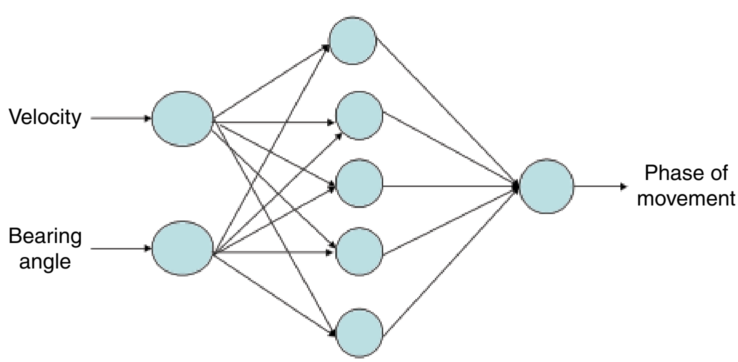


Figure 20.5 Neural network to predict phase of eye-gaze movement.

Applications

This section presents a set of case studies on using gaze-controlled interfaces in a variety of domains. The case studies cover desktop computing, automotive, and aviation environments. The first two studies did not use the target prediction algorithm whereas the third and fourth study did. The last study used both target expansion and multimodal technology.

Pointing and selection task

In this user study, we evaluated two multimodal eye-gaze tracking systems. These systems did not use any target prediction or expansion technology, and used a hardware switch for selection as we found it better than voice based selection in our previous study (Biswas & Langdon, 2015). We combined joystick-based and hand-movement-based pointing with eye-gaze tracking-based pointing.

Participants We collected data from 10 participants (age range 19 to 53, 5 male, 5 female), who did not have any visual, cognitive, or motor impairment. The participants were students and staff members of our university, and all of them took part in an eye-gaze tracking study once or twice, although they did not use a gaze-controlled interface regularly apart from taking part in user studies.

Material We conducted the study using a Windows 7 HP Pavilion computer (processor speed 2.5 GHz) and a 21" screen (435 mm × 325 mm) with 1600 × 1200 pixels resolution and a standard LogiTech computer mouse. We used a Tobii TX-2 (Tobii, 2013) eye gaze tracker along with the Tobii SDK. We also used a Leap motion controller (https://www.leapmotion.com/) and the US Air Force A10 Warthog HOTAS (http://www.thrustmaster.com/products/hotas-warthog) joystick and TARGET software to interface it with the operating system.

Design We tried to strike a balance between the complete natural interaction scenario of the input observer system (Evans & Wobbrock, 2012) and the controlled single-target task (Fitts, 1954; MacKenzie, Sellen, & Buxton, 1991) of traditional Fitts' law analysis. The task was like the ISO 9241 pointing task with multiple distractors on screen (Figure 20.6). Users were requested to click the button at the center of the screen and then the target button that appears with other distractors. The target button (white in Figure 20.6) can appear in the inner or outer ring at any random angle. The distractors were of same size as the target button, and the target and distractors were all square in shape.

During the study we used logging software that recorded cursor position and pupil size of participants in every 15 ms. The cursor logs were used to measure task-completion times and number of wrong selections while pupil diameters were analyzed to find a way to objectively measure cognitive load.

We compared the multimodal eye-gaze tracking technologies with respect to unimodal eye-gaze tracking.

Results In total we recorded more than 400 pointing tasks for each eye-gaze tracking-based system. We compared the point and selection times, TLX, and System Usability Scale (2014) scores among the unimodal nonadaptive eye-gaze tracking system

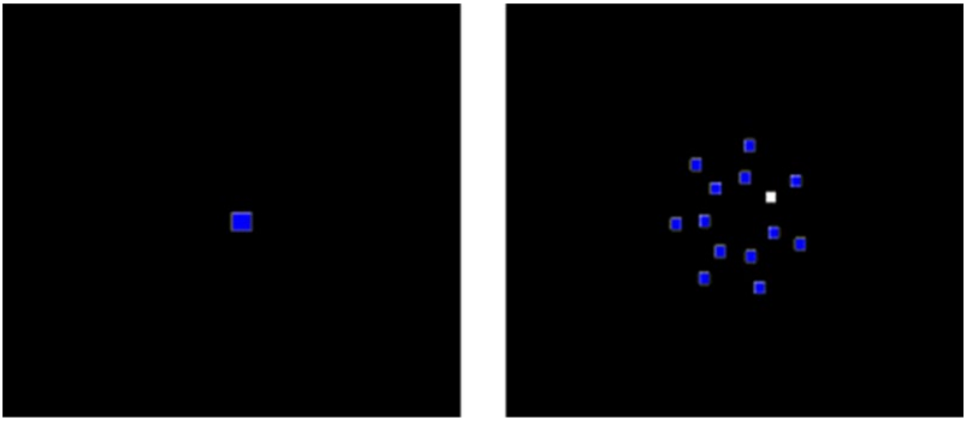


Figure 20.6 Multiple distractor task.

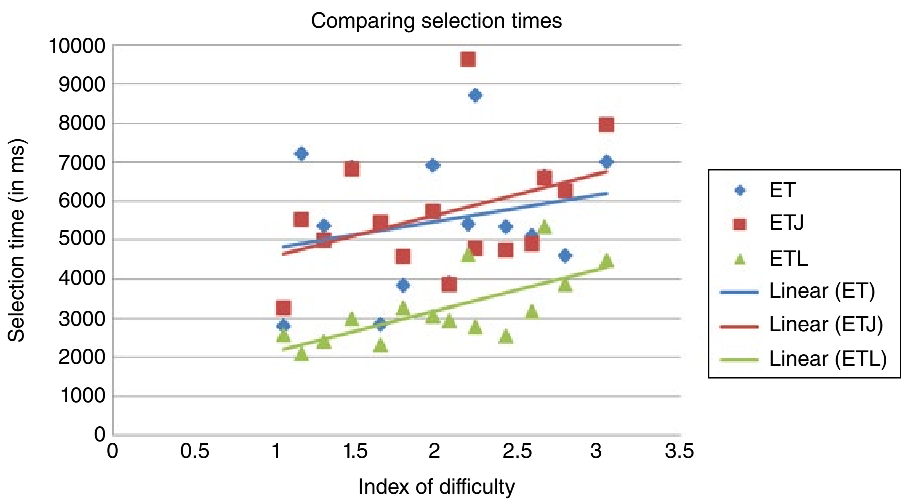


Figure 20.7 Comparing point and selection times between multimodal gaze-tracking systems.

and the multimodal eye-gaze tracking systems. Figure 20.7 plots the point and selection times with respect to IDs of targets. In a paired t-test we found that the point and selection time is significantly lower ($p < 0.01$) in the eye-gaze tracking with Leap motion system but the eye-gaze tracking with joystick is not significantly different from unimodal gaze-tracking system in terms of point and selection times.

The cognitive load in terms of TLX scores is found to be reduced in multimodal gaze-tracking system compared to the unimodal version and the difference is significant ($p < 0.05$) for the eye-gaze tracking with the Leap motion system. See Figure 20.8.

The number of wrong selections are found to be increased in eye-gaze tracking with joystick (ETJ) than the unimodal eye-gaze tracking-based system. The error was less than 1% as shown in Figure 20.9.

We also compared the subjective preferences of users between the multimodal gaze-tracking systems. Figure 20.10 plots the average SUS scores. It may be noted that a value of 68 in SUS signifies the system is usable and preferred by users. Users

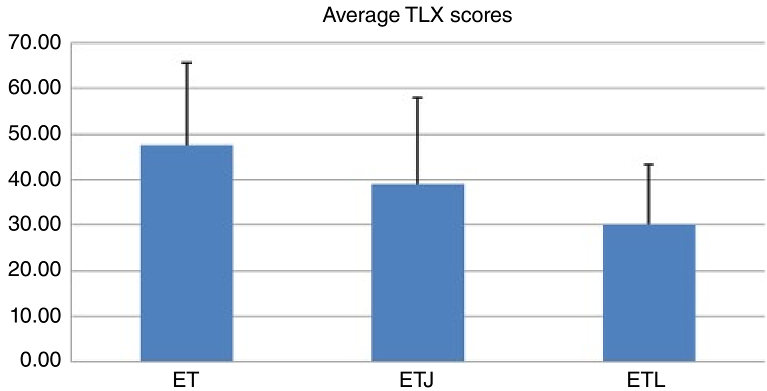


Figure 20.8 Comparing TLX scores between multimodal gaze-tracking systems.

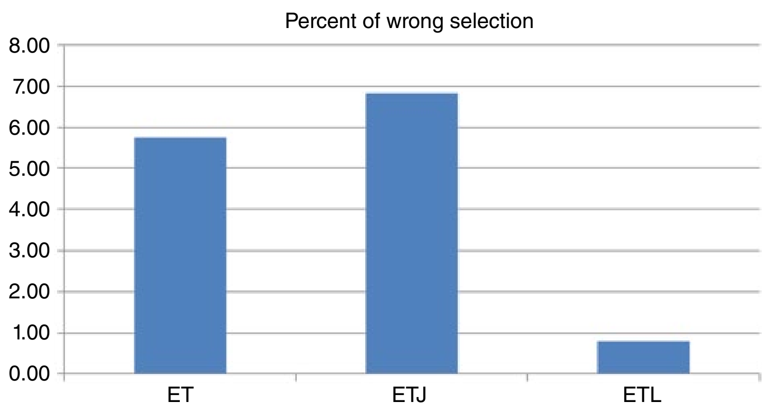


Figure 20.9 Comparing wrong selections between multimodal gaze-tracking systems.

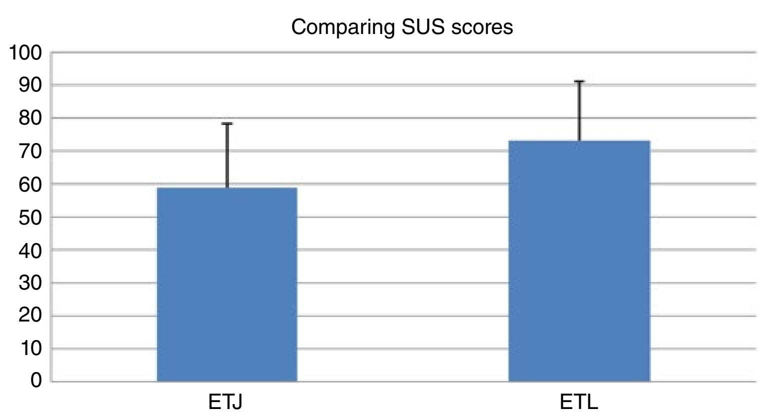


Figure 20.10 Comparing subjective preference between multimodal gaze-tracking systems.

preferred the multimodal gaze-tracking system with Leap motion to the multimodal gaze-tracking system with the joystick.

Discussion This study compared two different multimodal gaze-tracking techniques where we combined another pointing modality with a gaze-controlled interface. The addition of another pointing modality reduced the perceived cognitive load of users in terms of TLX score. Users could undertake a pointing and selection task faster in the eye-gaze tracking with the Leap motion-based system than the joystick-based system. The difference can be attributed to the particular model of joystick used in the study. Considering the application of the system in the aviation sector, we used a joystick attached to a throttle. The particular joystick uses a single button for moving the cursor in the X-Y plane as well as making a selection. Users often found it hard to make a selection using the joystick without moving the cursor in the X-Y plane and were also confused about switching between modalities using the joystick. The Leap motion-based system was less confusing as we used a separate hardware switch to make selection and users can easily switch between modalities just by putting their hand on top of the Leap motion and taking the hand away to use gaze tracking. However, there was latency in switching modality from gaze-tracking to hand-tracking as the Leap motion took a few milliseconds to detect hand movement, which occasionally increased pointing time, while there was no such latency in the joystick based system. In fact, the last user study described in this section used the joystick-based gaze-tracking system with target expansion technology, and was significantly faster than using only the joystick without gaze-tracking.

Map browsing

This study explored the use of eye-gaze tracking for large-scale spatial data processing. Unlike other studies described in this chapter, in this study, signals from the eye-gaze tracker were not used to control the onscreen pointer, rather the whole display. A technology demonstrator was developed involving Google Maps. Users could move the map and zoom in and out only using their eyes without involving hands. We developed the following interaction techniques involving the Google Maps:

- looking at the edge of the screen moves the map in opposite direction—for example if the user looks at the left edge of the screen, the map automatically scrolls towards right;
- if a user stares at a particular point in the map, that region zooms in;
- if the user blinks, the map is zoomed out.

The duration of staring and blinking could be configured although the following study used the same values for all participants. We also put on appropriate functions to distinguish between a conscious blink, unconscious blink, and signal loss from the tracker while the user looked away from the screen. A demonstration video of the system can be found at http://youtu.be/aJeiR_LZ1SE.

The following study compared users' cognitive load and subjective preference for the gaze-tracking interface with existing technology.

Participants We collected data from eight able-bodied participants (4 male, 4 female, ages ranging from 28 years to 35 years), who did not have any physical or cognitive impairments. They were all expert computer users and were familiar with the Google Maps interface.

Material We used an Acer Aspire E15 Laptop and a Tobii EyeX (Tobii, 2015) eye-gaze tracker and Tobii EyeX SDK. The laptop screen had a dimension of 34.5 cm × 19.5 cm and 1366 × 768 pixels resolution.

Design The study resembled a situation of searching visual stimuli from a spatial display. The participants were instructed to find four cities from a Google Maps display using eye-gaze tracking and the laptop touchpad. The order of using touchpad and eye-gaze tracker was randomized. The names of the cities were randomly chosen and were not visible in the default Google Maps interface. The cities surrounded the central location and were nearly same distance away from the central location.

The participants were instructed to find the cities and zoom on them when they found them. After the trial we instructed the participants to fill up questionnaires from NASA TLX, BRS, and SUS. We compared users' cognitive load and subjective preference for the touchpad and eye-gaze tracker.

Result We compared the BRS, TLX, and SUS scores among participants for the eye-gaze tracker and touchpad. All participants demonstrated through the BRS scores that they can complete the task in both conditions. Only one (P4) out of eight participants felt that the eye-gaze tracking condition should reduce the workload while the others felt the workload is either low or insignificant for both conditions.

Figure 20.11 shows the TLX scores for each participant whereas Figure 20.12 shows the SUS scores. In Figure 20.11, the bars correspond to the average score, and the error bar signifies the standard deviation.

Although the TLX and SUS scores are higher in the gaze-tracking condition than the touchpad the difference is not significant in a paired t-test. The difference is highest for TLX Mental Demand and Frustration.

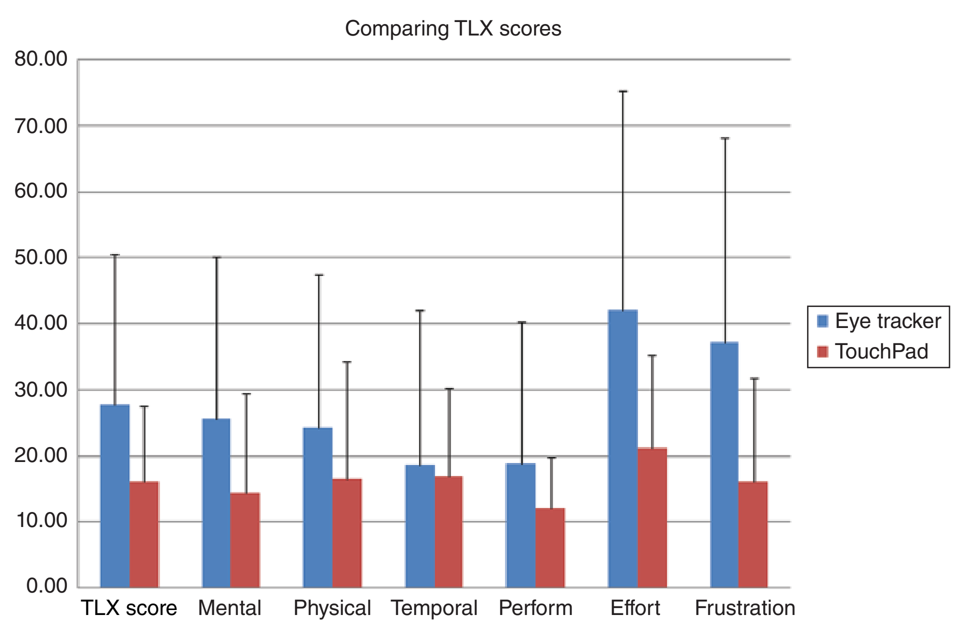


Figure 20.11 Comparing TLX scores for the map-browsing task.

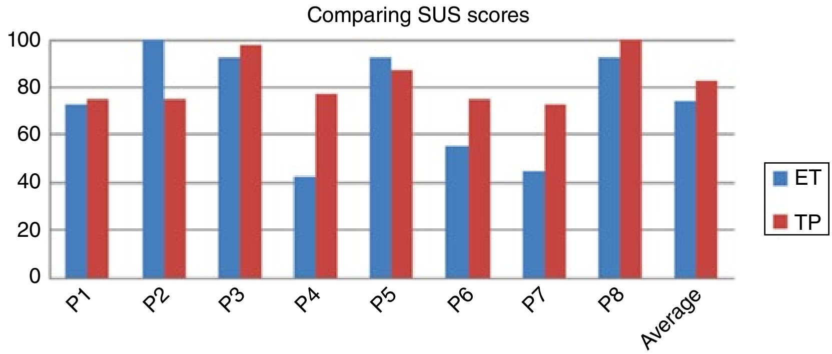


Figure 20.12 Comparing subjective preference in terms of SUS scores for the map-browsing task.

Discussion This study investigated the utility of the gaze-controlled interface for searching a visual stimulus in a large scale spatial display. A few prospective case studies may be found searching for a particular face in a surveillance video (a demonstration video can be seen at https://youtu.be/UjRoZbe9LAM) or investigating a particular molecular structure in a large topology, and so on. Our study shows that although users perceived higher cognitive load in the gaze-tracking interface than the conventional touchpad, the difference was not statistically significant and everyone could complete the task while they used the gaze-tracking system even first time. It is possible to integrate the multimodal systems described in the previous section with this map browsing system so that users can move or zoom the display using their eye-gaze and control an onscreen pointer using a joystick or Leap motion controller.

Electronic shopping

In this user trial we compared users' cognitive load and selection times between eye-gaze tracking and mouse for an online shopping task using the eShopping interface (Figure 20.13). This study uses target prediction technology with eye-gaze tracking. We collected data from participants who are not regular computer users. The study aimed to find how easy or difficult is it for users to perceive and perform with an eye-gaze tracking-based system in comparison to a mouse, which is still now the most commonly used computer input device. Vertegaal (2008) compared eye-gaze tracking and mouse-based interaction for pointing and clicking tasks and found that eye-gaze tracking with dwell-time based selection is faster than mouse but eye-gaze tracking also generated a higher error rate.

Participants We collected data from eight users (average age 57 years, 6 male, 2 female). Participants were interviewed about their prior experience of using computers and only allowed in the trial if they never used computer regularly before. A few users occasionally used computers but still did not consider themselves to be expert users.

Material We used a Windows 7 HP computer with a 54 cm × 33 cm monitor having 1920 × 1080 pixels resolution to record users' performance with the eShopping system. We used a Tobii TX2 eye-gaze tracker to record eye gaze. We used a Bezier curve

(Shirley & Marschner, 2009) based filtering algorithm to move the mouse cursor smoothly inside the screen. For eye gaze-tracking based interaction, the blank space button in a standard Logitech keyboard was used for the selecting target. A standard Logitech mouse was used to record mouse performance. We used the NASA TLX score sheet to measure cognitive load.

Design The users were instructed to buy a few items using the eShopping interface (Figure 20.13) using the mouse and then the eye-gaze tracker. The mouse-based interaction did not involve a target prediction system while the eye-gaze tracking-based system had the target prediction on. After repeating the process a few times, they were instructed to fill up the TLX score sheet. The order of input options (mouse and eye-gaze tracker) was randomized to minimize order effect. The process of buying an item involved the following steps:

1 Pointing and clicking on one of the combo boxes on top (Figure 20.13).
2 Pointing and clicking on the button having the desired item (like camera, computer etc.—see Figure 20.13). On clicking a button, the interface shows a list of cameras, computers, and so on.
3 Pointing and clicking on the button having the desired product like a particular computer brand or a particular book.
4 Repeating the above steps to add more items to the shopping cart.
5 Pointing and clicking on the "check out" button at the right side of the screen (Figure 20.13).
6 Repeating the whole procedure (steps 1–5) two to three times using both mouse and eye-gaze tracker.

Results All eight users could undertake the trial and completed the task. The button selection time was measured as the difference in time between two button selections or the time difference between selection of a combo box and next button press. The time involves pointing to the target and selecting it. The button selection time was

Figure 20.13 Electronic shopping interface.

Figure 20.14 Comparing average selection time for electronic shopping task.

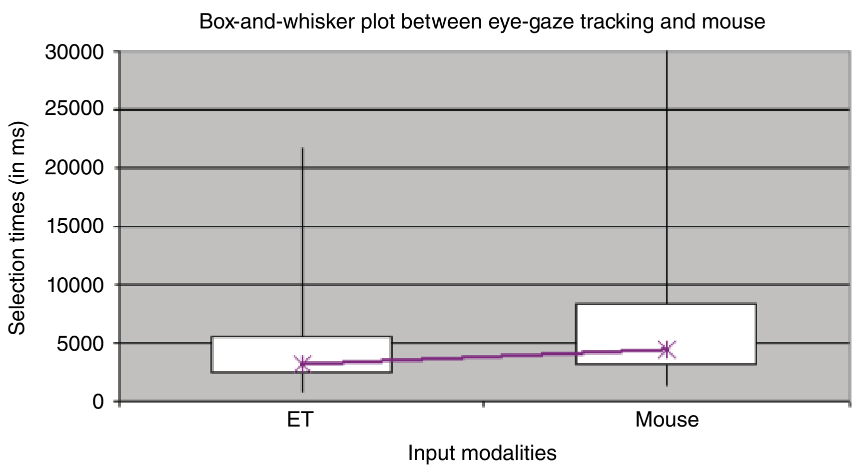


Figure 20.15 Comparing box and whisker plots on selection time for electronic shopping task.

significantly less for the eye-gaze tracking-based system than for the mouse (Figure 20.14 shows average and standard deviation and Figure 20.15 shows median and quartiles) in a Wilcoxon signed-rank test ($Z = -2.84$, $p < 0.01$, $r = -0.33$). In the experimental set up, we defined error or wrong selection as follows:

- users selecting same item twice consecutively;
- users selecting "remove last item" button;
- users selecting "clear all" button.

We found users committed four wrong selections among 93 selections for the eye-gaze tracking system and one wrong selection among 79 selections using the mouse. The error rate is below 5% in both cases.

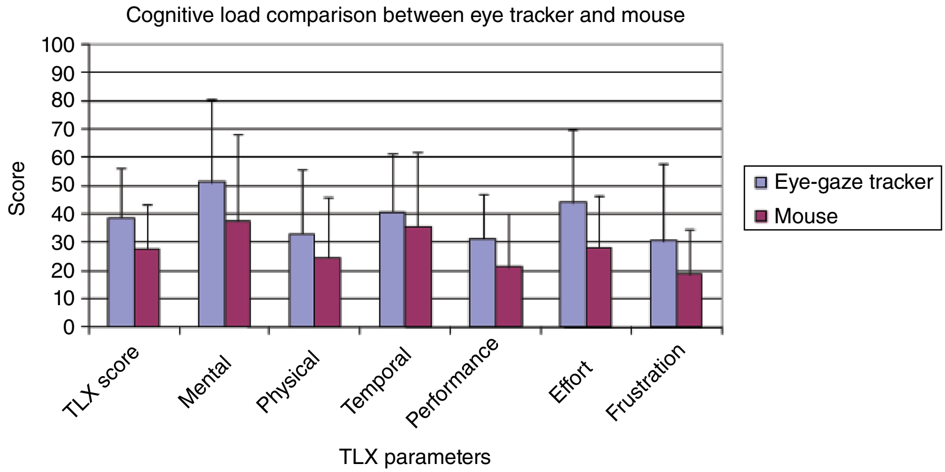


Figure 20.16 Comparing TLX scores for electronic shopping application.

Figure 20.16 shows the cognitive load in terms of NASA TLX scores. The columns correspond to average scores while the Y error bars signify standard deviation. Users scored higher TLX scores for the eye-gaze tracker (mean 38.48, stdev 17.85) than the mouse (mean 27.66, standard deviation 15.67), although the difference was not significant in a paired two-tailed t-test.

Discussion This study demonstrates that for an easy-to-use interface, novice users can complete tasks quicker using an eye-gaze tracker than a mouse, although the eye-gaze tracker tends to produce more cognitive load than the mouse. It may be noted none of these users had used an eye-gaze tracker before although six of them used a mouse before. We recorded only four occasions where users took more than 10 s to select a button among 93 correct selections. The average button selection time was 4.3 s.

Automotive dashboard control

Kern, Mahr, Castgronovo, Schmidt, and Müller (2010) and Poitschke, Laquai, Stamboliev, and Rigoll (2011) reported user studies involving simulated driving tasks while comparing an eye-gaze controlled interface with a traditional one. The present study explored the possibility of a gaze-control interface for operating a dashboard in an automotive environment. In particular, we evaluated the effect of two different track conditions on drivers' performance with an eye-gaze tracking interface. Kern et al. (2010) and Poitschke et al. (2011) reported user studies involving simulated driving tasks while comparing an eye-gaze controlled interface with a traditional touch-screen control. We took forward that work with a low-cost eye-gaze tracker and an intelligent target prediction algorithm that can reduce pointing time. A demonstration video of the system can be found at http://youtu.be/lmYZcnwzEbU

Participants We collected data from 12 participants (age range 19 to 27, 10 male, 2 female). All participants were university students, and none of them regularly drove cars. Eight participants had driving licenses although the qualities of driving tests were quite different for them. However, all participants were expert users of the driving simulator and used to drive cars in the simulator.

Design We designed the test to evaluate the effect of an eye-gaze controlled secondary task on the primary driving task with participants with varying level of driving skills. The primary task involved driving a car in the left lane without veering off from the lane. We used two different track conditions—a simple track consisting of four turns and a complex track consisting of 20 turns. There was no other traffic on the road, and drivers were instructed to drive safely without veering off the driving lane and simultaneously operating the car dashboard using their eye gaze. The secondary task was initiated through an auditory cue. It mimicked a car dashboard (Figure 20.17) and participants were instructed to press a button on it after hearing the auditory cue (Figure 20.18). The auditory cue was set to appear between 5

Figure 20.17 Secondary task in automotive study.

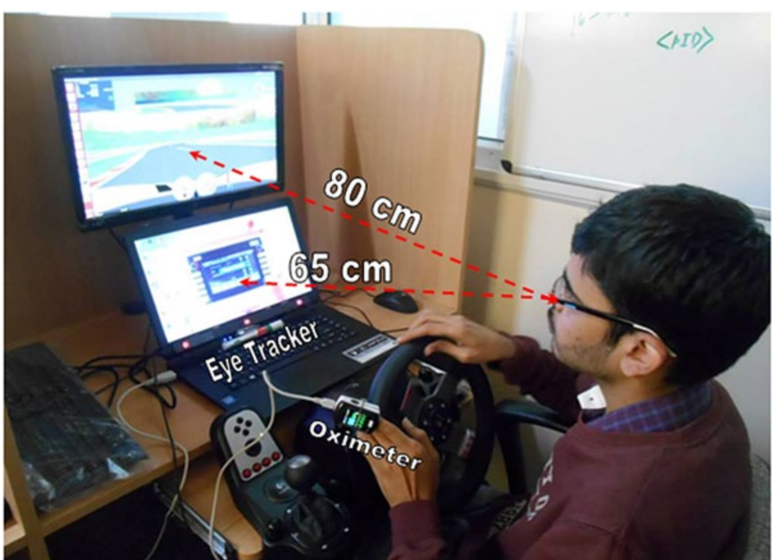


Figure 20.18 Experiment design for automotive task.

and 7 s intervals. The target button was randomly selected in the car dashboard. The pointing was undertaken through the eye-gaze of users using an intelligent eye-gaze tracking algorithm (Biswas & Langdon, 2015) and selection was done through a hardware button on the steering wheel.

The study (Figure 20.18) was a 2×2 factorial design where the independent variables were:

- Track condition:
  - simple;
  - complex.
- Presence of secondary task:
- driving without secondary task;
- driving with secondary task.

The dependent variables were:

- task completion time;
- average deviation from the center of the road;
- number of correct selections in gaze-controlled interface.

We also measured drivers' cognitive load in terms of pulse rate using an Oximeter (http://www.nonin.com/What-is-Pulse-Oximetry) and NASA TLX scores.

Material We used Logitech driving simulator hardware and Torque© car simulation software. The hardware was set as an automatic transmission car. We used a Tobii EyeX eye-gaze tracker and EyeX SDK for the gaze-controlled interface. The primary task was run on a Linux desktop while the secondary task was conducted on a Windows 8 Laptop. The Laptop screen had a dimension of 34.5 cm × 19.5 cm with screen resolution of 1368 × 800 pixels.

Procedure Initially participants were briefed about the procedure and trained to use the driving simulator and the gaze-controlled interface. Then they undertook the trial in random order of track conditions. After completion of each condition, they filled up the TLX sheet based on their toughest experience during the trial.

We used logging software that recorded the trajectory of the car with timestamps from the driving simulator and cursor and eye-gaze movements from the secondary task. We also recorded participants' pulse rate from the oximeter with the timestamp.

Results We found a statistically significant correlation between number of correct selections in the secondary task and average velocity of the car (Figure 20.19, $r = -0.46$, $p < 0.05$). Drivers could make a significantly higher number ($t(1,21) = -2.2$, $p < 0.05$) of correct selections using eye-gaze control while they were driving in the complex track than in the simple track (Figure 20.20). In a repeated measure ANOVA, we found:

- significant main effect of track condition on
  - task completion time $F(1, 11) = 88.24$, $p < 0.01$, $\eta^2 = 0.89$;
  - deviation from driving lane $F(1, 11) = 6.51$, $p < 0.05$, $\eta^2 = 0.37$;
  - TLX score $F(1, 11) = 14.58$, $p < 0.01$, $\eta^2 = 0.57$.

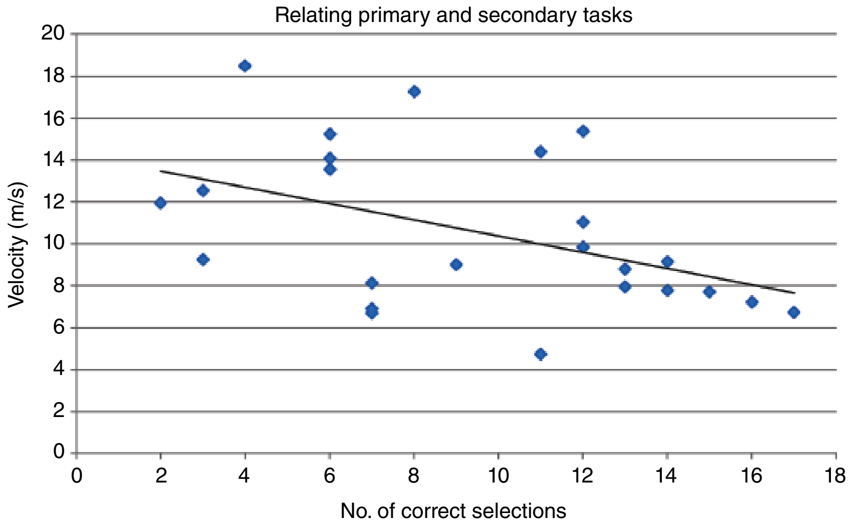


Figure 20.19 Average driving velocity is correlated with number of correct selections in secondary task.

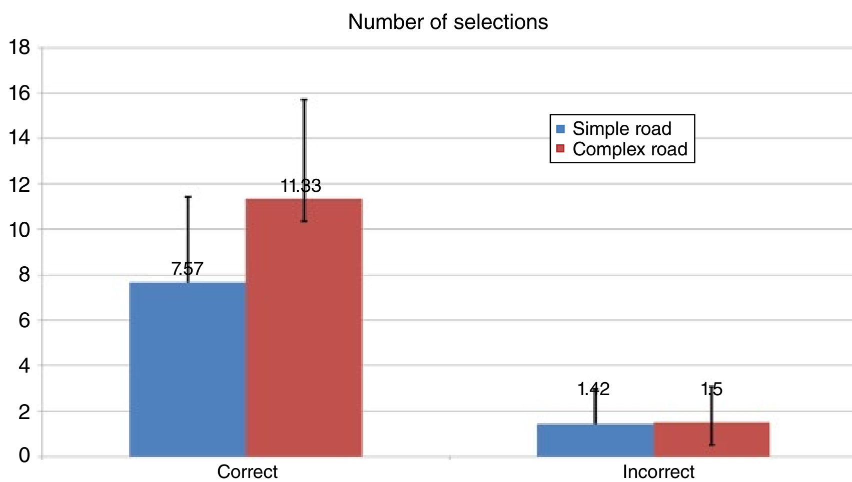


Figure 20.20 Number of selections in secondary tasks in different road conditions.

- Significant main effect of presence of secondary task on:
  - task completion time $F(1, 11) = 22.07$, $p < 0.01$, $\eta^2 = 0.67$;
  - deviation from driving lane $F(1, 11) = 13.69$, $p < 0.01$, $\eta^2 = 0.55$;
  - TLX score $F(1, 11) = 23.01$, $p < 0.01$, $\eta^2 = 0.68$.

The interaction effects were not significant for any variable at $p < 0.05$. It may be noted that the presence of a secondary task had a bigger effect on deviation from the driving lane and TLX scores than the track condition while the track condition had a bigger effect on task completion time than the presence of a secondary task. The result indicates that users adjusted their speed of driving based on road condition and

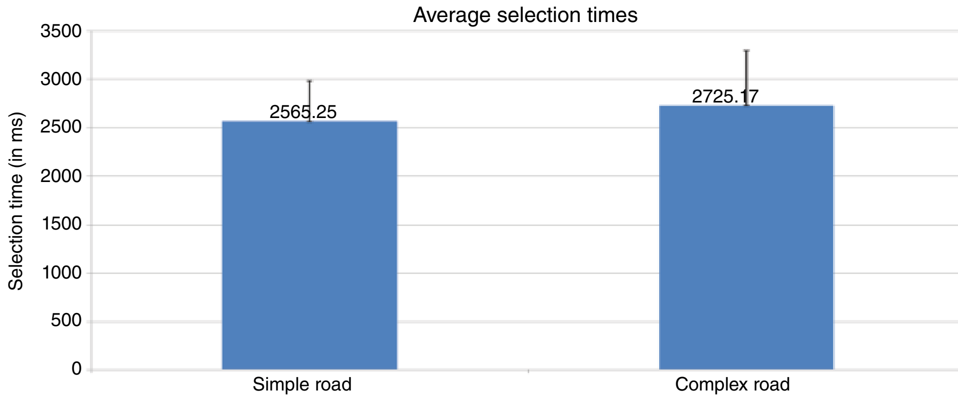


Figure 20.21 Average selection times in gaze control interfaces for two different road conditions.

drove slower in the complex track. As they drove slowly, they could undertake more pointing and selection tasks in the complex track than the simple track. However, when they were involved in a secondary task, they tended to deviate from driving lane more often than without any secondary task.

We measured the time difference between the instances of an auditory cue and selection of a target button in the gaze controlled secondary task interface. This time difference is equal to the pointing and selection time of the target button using eye-gaze. Use of the intelligent eye gaze tracking reduced the pointing and selection time to 2.5 s on average even for novice users who had not used a gaze-control interface earlier (Figure 20.19). The difference in selection times (Figure 20.21) for two different track conditions were not significant at $p < 0.05$.

In summary, we concluded:

- complexity and the presence of dual tasks significantly increases cognitive load and task-completion times;
- performance with a secondary task is significantly related to the velocity of the car—in a complex road condition, users drove slowly and performed better with a secondary task than in a simple road condition;
- with the present state of eye-gaze trackers, users needed approximately 2.5 s for pointing and selection.

Aircraft cockpit control

This task explored the possibility of using the multimodal adaptive eye-gaze tracking system in the cockpit of a combat aircraft. We aimed to augment the existing hands-on-throttle-and-stick (HOTAS) joystick with the eye-gaze tracking system. We compared the multimodal eye-gaze tracking system with a HOTAS-based joystick. The task involved participants to check five targets in a simulated multifunction display, which had same dimension as the original one in the Eurofighter Typhoon aircraft. The following sections describe the study in details.

Participants We recruited eight young able-bodied participants (5 male, 3 female, average age 31.2 years).

Material We conducted the study using a Windows 7 HP Pavilion computer (processor speed 2.5 GHz) and a 21" screen (435 mm × 325 mm) with 1600 × 1200 pixels resolution and a standard LogiTech computer mouse. We used a Tobii TX-2 (Tobii, 2013) eye-gaze tracker along with Tobii SDK. We used the USAF A10 Warthog HOTAS (http://www.thrustmaster.com/products/hotas-warthog) and the TARGET software to interface it with the operating system.

Design The task involved selecting a set of five targets in a simulated multifunction display (Figure 20.22) and resetting the display after clicking on them. Participants used the multimodal intelligent eye-gaze tracking (the one that combines both eye-gaze tracking and joystick-based pointing) and HOTAS-based joystick. We used the target prediction technology for both devices. We also investigated cognitive load as well as subjective preferences of users and collected TLX scores and System Usability Scale (Brooke, 1996) scores for each modality.

Procedure Participants were initially briefed about the task. The task involved browsing through a menu tree to make a target visible and then clicking on the target. The target would have appeared anywhere on the screen. Target width and distances were in the same range as in the previous study. A trial involved the selection of five targets constituting at least 20 pointing and selection tasks. Each participant undertook the trial twice using each modality. The order of modalities was randomized. After completion of the trial in one modality, participants filled up TLX and SUS score sheets based on their average performance.

Figure 20.22 Simulated multifunction display for aviation study.

Results We initially compared target selection times for both input modalities. The selection time was calculated from the moment of selecting a button to the moment of selecting the next button. We ignored the first selection of each trial as the system started the logging procedure after it. A device (2) × session (2) ANOVA found a significant effect of device ($F(1,153) = 26.07$, $p < 0.0001$, $\eta^2 = 0.15$) and session device ($F(1,153) = 4.99$, $p < 0.05$, $\eta^2 = 0.03$), although the interaction effect was not significant (Figure 20.23).

The TLX score (Figure 20.24) was lower for eye-gaze tracking although not significantly different in a t-test from a joystick. However, the TLX Frustration score was significantly lower ($p < 0.05$) with eye-gaze tracking than with a joystick.

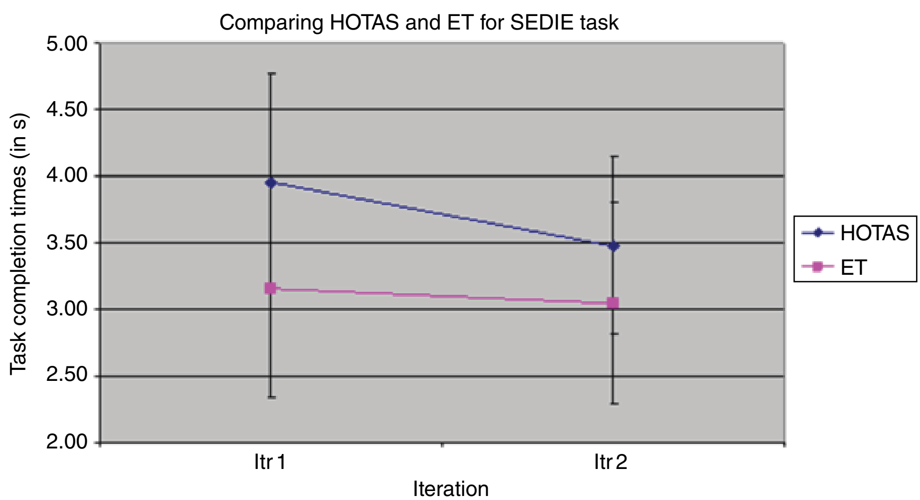


Figure 20.23 Comparing selection times for military task.

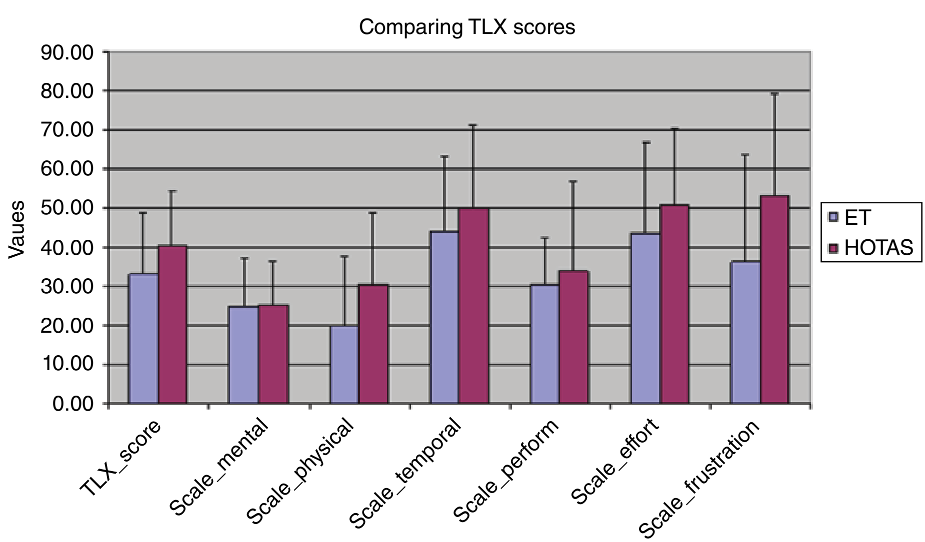


Figure 20.24 Comparing TLX scores for aviation task.

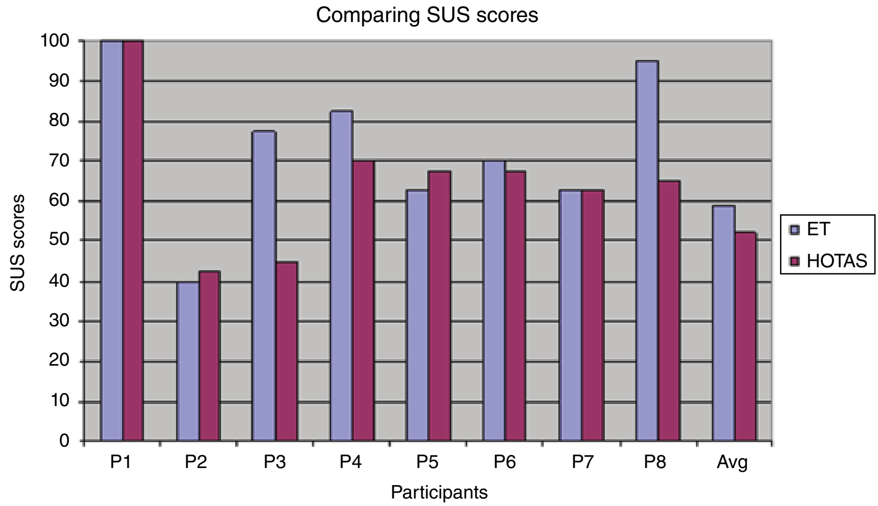


Figure 20.25 Comparing subjective preference in terms of SUS scores for aviation task.

Figure 20.25 summarizes results from the SUS questionnaire. SUS employs a five-point scale from strongly agree to strongly disagree. We compared the number of people who agreed (either strongly or only agreed) and disagreed (either strongly or only disagreed) for each device. We found that more users preferred the eye-gaze tracking-based system to the joystick and felt confident to use the eye-gaze tracking-based system.

Discussion This study further confirms that users can undertake trials for a realistic military task using the intelligent eye-gaze tracking system faster than the existing HOTAS-based joystick. Users also perceived less cognitive load for the eye-gaze tracking system than the joystick, which was also reflected in their subjective preferences in terms of the SUS scores. This study demonstrates another example of augmenting existing interaction devices with eye-gaze tracking. The joystick can be configured based on different multifunction displays but moving the pointer for both large-scale movements and precise homing movements may be challenging not only for novice users but also for expert users under a high-workload situation. Using eye-gaze tracking to move the pointer near the target or even selecting the appropriate display and then switching to the usual joystick input is a promising solution, as demonstrated in our study.

Summary

The eye-gaze controlled interface was mainly explored for people with severe disabilities and, recently, for improving computer gaming experience. The set of case studies in this chapter aims to extend the scope of a gaze-controlled interface. The case studies start with basic pointing and selection tasks and then move to applications for map browsing, computer novice users, automotive, and military aviation environments. It may be noted that our participants did not use a gaze-controlled interface before taking part in these studies but still the gaze-controlled interface either improved (in case of computer novice users and military aviation case study) or did not significantly slow down speed of interaction. We emphasize that gaze-controlled interfaces should not only be confined to specialized applications but can be extended to a plethora of domains, even for able-bodied or so-called average users.

Conclusions

Farrell and Zhai (2005) noted that "humans use their eyes naturally as perceptive, not manipulative, body parts. Eye movement is often outside conscious thought, and it can be stressful to carefully guide eye movement as required to accurately use these target selection systems." However, it may also be noted that interaction with any graphical user interface involves visual search and we can leverage this visual search to select targets, too. As Farrell noted, if the process needs "careful guidance" of eye movements it would be stressful but in this chapter we proposed a target prediction system and integrating other modalities that can reduce pointing and selection times, also requiring less precise control of conscious eye-gaze movements. There are also situations where existing pointing devices are not an optimal choice or are dangerous to use. For example for people with severe motor impairment or operators in aviation and automotive environments cannot easily use a mouse or touchpad like their able-bodied counterparts or like those in desktop computing situations. The combination of eye-gaze tracking with scanning will be useful assistive technology and the user study above demonstrates an intelligent eye-gaze controlled interface can be useful for able-bodied novice computer users as well. Considering the case of situational impairment, aviation and automotive user interfaces do not require continuous manipulation of an onscreen pointer like graphical user interfaces in desktop computing. Unless a particular interaction is very familiar to the driver (like reaching for the gearbox while driving), he has to glance at the user interface. Accurate gaze-tracking with target prediction technology can leverage this glance for pointing. Additionally the same eye-gaze tracker can be used to detect cognitive load or distraction from driving or piloting, which in fact can increase the safety of driving or flying.

References

Bates, R. (1999). Multimodal eye-based interaction for zoomed target selection on a standard graphical user interface. In *Proceedings of INTERACT*. London: British Computer Society.

Biswas, P., & Langdon, P. (2011). A new input system for disabled users involving eye gaze tracker and scanning interface. *Journal of Assistive Technologies, 5*(2), 58–67.

Biswas, P., & Langdon, P. (2015). Multimodal intelligent eye-gaze tracking system. *International Journal of Human-Computer Interaction, 31*(4), 277–294.

Brooke, J. (1996). SUS: a "quick and dirty" usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester, & A. L. McClelland (Eds.), *Usability evaluation in industry*. London: Taylor & Francis.

Evans, A. C., & Wobbrock, J. O. (2012). Taming wild behavior: The input observer for obtaining text entry and mouse pointing measures from everyday computer use. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '12)* (pp. 1947–1956). New York, NY: ACM.

Farrell, S., & Zhai, S.(2005). *US Patent No. 20050047629A1*. System and method for selectively expanding or contracting a portion of a display using eye-gaze tracking. Washington, DC: U.S. Patent and Trademark Office.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology, 47*, 381–391.

Kern, D., Mahr, A., Castronovo, S., Schmidt, A., & Müller, C. (2010). Making use of drivers' glances onto the screen for explicit gaze-based interaction. In *Proceedings of the Second International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 110–116). New York, NY: ACM.

Lank, E., Cheng, Y. N., & Ruiz, J. (2007). Endpoint prediction using motion kinematics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)* (pp. 637–646). New York, NY: ACM.

MacKenzie, I. S., Sellen, A., & Buxton, W. (1991). A comparison of input devices in elemental pointing and dragging tasks. In *Proceedings of the CHI '91 Conference on Human Factors in Computing Systems* (pp. 161–166). New York, NY: ACM.

Poitschke, T., Laquai, F., Stamboliev, S., & Rigoll, G. (2011). Gaze-based interaction on multiple displays in an automotive environment. In *IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 543–548). doi:10.1109/ICSMC.2011.6083740

Shirley, P., & Marschner, S. (2009). *Fundamentals of computer graphics.* Boca Raton, FL: CRC Press.

Tobii. (2015). Tobii EyeX Eye Tracker, Available at: http://www.tobii.com/xperience/

Tobii. (2013). Tobii TX2 Eye Tracker, Available at: http://www.tobii.com/en/eye-tracking-research/global/products/hardware/tobii-x60x120-eye-tracker/.

Vertegaal, R. (2008). A Fitts' law comparison of eye tracking and manual input in the selection of visual targets. In *Proceedings of the International Conference of Multimodal Interaction* (pp. 241–248). New York, NY: ACM.

Zandera T. O., Gaertnera M., Kothea C. & Vilimek, R. (2010). Combining eye gaze input with a brain–computer interface for touchless human–computer interaction. *International Journal of Human-Computer Interaction, 27*(1), 38–51. doi:10.1080/10447318.2011.535752

Zhai, S., Morimoto, C., & Ihde, S.(1999). Manual and gaze input cascaded (MAGIC) pointing. *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (pp. 246 -–253). New York, NY: ACM.

Ziebart, B., Dey, A., & Bagnell, J. A. (2012). Probabilistic pointing target prediction via inverse optimal control. In *Proceedings of the 2012 ACM International Conference on Intelligent User Interfaces (IUI '12)* (pp. 1–10). New York, NY: ACM.

21

Corneal Imaging

Christian Nitschke and Atsushi Nakazawa

Preliminary Comments

Since *corneal imaging* is (and will probably remain in the foreseeable future) a niche topic, questions like these may occur: What is this about? Is this of interest to me? Should I read it? How should I read it? This chapter therefore conveys background information to help potential readers find their individual answers and access the information efficiently.

The aim of this chapter is to provide a comprehensive introduction and overview, which covers related topics and their implications. Starting in the early 2000s, the field has developed from a not-so-serious investigation of what may be possible, to proof-of-concept work, to robust practical implementations and realistic applications. Apart from its growing relevance, it is a highly interdisciplinary topic with relation to a diverse range of fields, shown by the following list of problems that appear in this chapter, which is not exhaustive:

- human–computer interaction (HCI) (human sensing; face tracking; eye tracking; point-of-gaze tracking; human intent recognition; assistance systems);
- computer vision (face / facial feature tracking; face reconstruction; catadioptric imaging; image enhancement; image registration; structured light; illumination modeling; scene reconstruction; scene / context recognition);
- computer graphics (eye / face modeling; scene illumination and relighting);
- forensics (scene context, location, and time estimation of face image acquisition; forgery detection in visual media);
- augmented / virtual reality (calibration of head-mounted / head-up displays);
- biometrics (iris recognition; face recognition);
- anatomy and medicine (eye anatomy; computational modeling of appearance, shape, and dynamics of eye structures; high-quality eye imaging);
- psychology (cognitive load estimation from pupillary response; stimulus-response analysis from scene and face imagery; human ability to recognize faces, emotions, and interest from low-quality imagery).

The Wiley Handbook of Human Computer Interaction, Volume 1, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

For several years, the authors worked on eye-imaging- and eye-modeling-related topics, developing hardware, methods, and applications for robust nonintrusive human sensing. This chapter is a major extension and revision of the overviews in Nitschke (2011) and Nitschke, Nakazawa, and Takemura (2013b), including recent developments and improved presentation to target a wider audience. A large part stems from the authors' own work with care taken specifically to integrate achievements from the whole field and provide a balanced and objective overview.

The content is organized by context (rather than field) to allow for a holistic view. While the common thread is a high-level discussion, it includes extensive details and references to the relevant literature to allow for insights on a specific topic. In summary, this chapter targets a diverse audience with different backgrounds, levels of experience, and theoretical and practical interest.

Introduction

Our eyes are one of the most important sensory organs, allowing the exploration, analysis, perception of, and interaction with the visual information content of the physical world. Therefore, the eye and its movements provide a key contribution to interpreting and understanding a person's wishes, needs, tasks, cognitive processes, affective states and interpersonal relations. The unique geometric and photometric properties of the eyes provide important visual cues for obtaining face-related information; their unique appearance is exploited in biometrics and computer graphics. Image-based eye analysis allows for nonintrusive measurement of eye structures and visual acuity in optometry, and diagnosis of diseases and disorders of the visual system in ophthalmology.

Corneal imaging

The cornea is the transparent protective and optical outer layer of the eye that covers the iris and accounts for the majority of the eye's optical power. While light arriving at its surface mainly refracts and enters the eye, a small part reflects back into the environment and can be noticed when looking at a person's eye (Figure 21.1). A known shape and pose for the cornea enables the cornea to be modeled—this uses camera geometry as a catadioptric imaging system[1] to recover the environmental illumination from an image of the eye, and applies it to various tasks in computer vision and computer graphics. This approach is commonly referred to as *corneal imaging* (Nishino & Nayar, 2006; Nitschke, 2011; Nitschke et al., 2013b). Analyzing and exploiting such corneal reflections from eye images provides a direct relationship between the environment (scene panorama / model) and an observer (face / eye pose), which allows for a wide range of intriguing applications and novel solution approaches to existing problems.

Applications

Modeling corneal reflections allows the situation under which a person is photographed (surveillance, forensics) to be determined, allows the calculation of a person's field of view and point of gaze (PoG) (human sensing, HCI), and allows a higher level

[1] A catadioptric system is an optical system that combines refraction and reflection, commonly achieved via lenses (dioptrics) and curved mirrors (catoptrics).

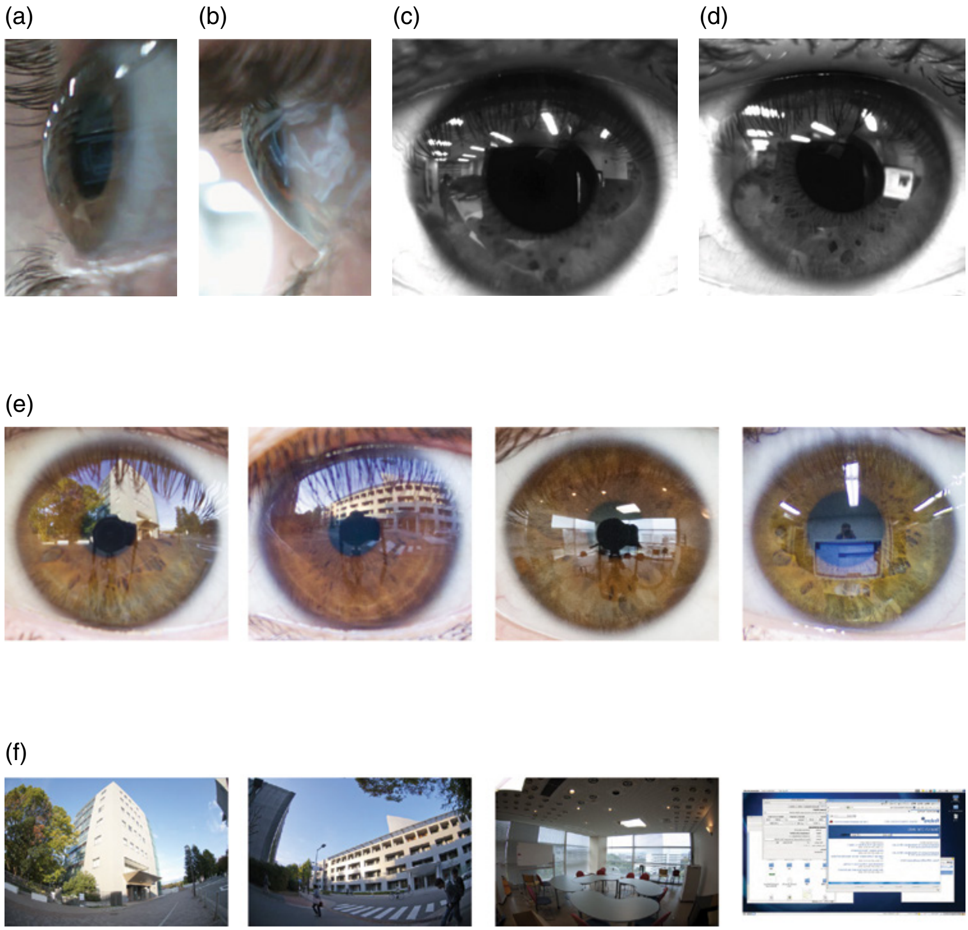


Figure 21.1 Corneal reflections. (a) View from the side onto the transparent reflective cornea. (b) A close view reveals the corneal limbus as the surface shape discontinuity, where the cornea dissolves into the white sclera. (c) The reflected office environment is clearly visible in the eye image. Diffuse iris reflections are superimposed with specular corneal reflections. (d) Focus on iris texture instead of corneal reflections. (e),(f) Examples of corneal images and corresponding (fisheye) images for different scenes.

analysis of stimulus and response (psychology) to be performed, which may enable novel insights and interface concepts: Backes, Chen, Dürmuth, Lensch, and Welk (2009) describe an image-based eavesdropping technique for recovering reflected content from computer displays at faraway locations. For this purpose, they analyze the point spread function (PSF) of the corneal reflection system and introduce a nonblind deconvolution method to compensate for defocus and motion blur. Nishino and Nayar (2006) provide the first comprehensive analysis of the visual information about the environment that is embedded within an image of the human eye. Their seminal study formalizes the combination of camera and corneal reflector as a catadioptric imaging system, derives its imaging characteristics, and describes its geometric calibration through eye modeling and pose estimation. The recovered environment map of incident illumination at the eye allows for a number of interesting applications, such as the computation of a panorama or a subject's view of the scene. The latter is further improved to achieve nonintrusive calibration-free eye-gaze tracking (EGT) in arbitrary dynamic environments (Nakazawa & Nitschke, 2012; Nakazawa, Nitschke, & Nishida, 2016; Nitschke, Nakazawa, & Nishida, 2013a; Nitschke, 2011; Takemura, Kimura, & Suda, 2014a; Takemura, Yamakawa, Takamatsu, & Ogasawara, 2014a, b). Interestingly, corneal imaging even made it into science-fiction works, where they showcase several exciting applications (Motion Pictures, 2004). For example, by applying (some fancy) corneal reflection extraction to a surveillance video footage that shows a persons's facial region, they obtain detailed information about the surrounding scene that the person in the video looks at, revealing a criminal's face or printed fabric patterns. Inspired by this, Nitschke and Nakazawa (2012) describe a practical super-resolution strategy for corneal imaging that allows the simplification of image acquisition and the improvement of image quality by restoring lost details.

Beside extracting visual information from the scene, the recovered environment map (encoding light intensity and direction) can be further applied to various computational tasks in vision and graphics such as face modeling / relighting, image forensics and biometrics: Tsumura, Dang, Makino, and Miyake (2003) are the first to recover the direction of environmental illumination from specular highlights in the eye and apply it to face reconstruction by photometric stereo and face relighting. Nishino and Nayar (2006) apply the complete environment map for improved face reconstruction and relighting (Nishino & Nayar, 2004), and for illumination normalization in face recognition (Nishino, Belhumeur, & Nayar, 2005). Johnson and Farid (2007) describe a method to reveal digital forgeries, where images are composed with people photographed at different places or times by analyzing inconsistencies in the illumination distribution from the corneal reflections of all subjects. The method can be applied with any arbitrary photo as it automatically estimates the required internal camera parameters from the perspective distortion of the iris contour.

The two eyes, captured in a single face image, form a catadioptric stereo system that enables reconstruction of a simple 3D scene structure (Nishino & Nayar, 2006). It has been shown that the pose of a planar computer monitor or projection screen can be estimated from reflected point (Nitschke, Nakazawa, & Takemura, 2011a) or line features (Schnieders, Fu, & Wong, 2010) under multiple eye poses. The combination of known poses for scene, eyes, and camera—and in the case of screens and active illumination, the known illumination information—facilitates a large number of sensing and interactive applications in various fields. This includes surveillance and forensics (Backes et al., 2009), computer graphics and vision (Nakazawa et al., 2016; Nishino & Nayar, 2006; Tsumura et al., 2003), HCI (Hansen & Ji, 2010; Nakazawa & Nitschke, 2012; Nakazawa et al., 2016; Schnieders et al., 2010), augmented and virtual reality (AR/VR) (Itoh & Klinker, 2014; Plopski, 2016; Plopski et al., 2015a), and medicine (Pamplona, Mohan, Oliveira, & Raskar, 2010; Pamplona et al., 2011).

Limitations, challenges, and potential

While corneal imaging enables catadioptric vision (Yagi, 1999) with a large number of potential applications, the cornea is not a perfect mirror and suffers from several issues. These include (a) a low resolution from the small size and large field of view of the corneal mirror, (b) a low contrast as the reflectivity of the cornea is less than 1% (Nishino & Nayar, 2004), (c) a contamination with iris texture reflections (Wang, Lin, Ye, & Gu, 2008) as the cornea is transparent, and (d) distortions from an unknown individual shape that is not handled by simple eye models. The low reflectivity further makes image acquisition challenging, as it requires a long exposure and an open aperture, which causes motion and defocus blur. Corneal reflection analysis benefits from theories and techniques in catadioptric imaging (Sturm, Ramalingam, Tardif, Gasparini, & Barreto, 2011), but at the same time demands for specialized algorithms that handle and exploit the unique properties of the eye. Regarding the complexity of the problem, the majority of algorithms either completely refrain from geometric modeling or apply a substantial simplification, which often sacrifices quality, usability, and applicability (Nitschke et al., 2013b).

Corneal reflection analysis has a great potential, but present advancement is limited by several factors that need to be handled. These include (a) the low quality of corneal images (resolution, reflectivity, iris texture), (b) the unknown characteristics of the individual eye (shape, appearance), (c) the establishment of reliable automatic modeling techniques, and (d) the integration of multiple eye and scene images (correspondence matching) with available domain knowledge. Regarding the limitations and challenges, corneal imaging will not replace high-quality catadioptric systems, but can enable information to be gathered when only face/eye images are available or when the relationship between a person and the environment is concerned. An advantage is the minimum requirement of only a single face image that provides a distinct view of the scene from each eye. This naturally allows for ad hoc application to dynamic and real-time scenarios, which is commonly not achieved with standard approaches.

Outlook

The remainder of this chapter provides a comprehensive introduction on corneal imaging, covering four major parts: (a) An introduction to human eye modeling in computer vision, computer graphics, and HCI. This covers the anatomic background, the geometric modeling, and the image tracking and 3D pose estimation. (b) An overview of corneal reflection modeling and processing techniques. This comprises basic modeling and more advanced image processing. (c) A review of applications in various fields showcases the wide range and implications of the topic, and summarizes the state of the art. (d) Finally, a discussion of promising future directions highlights strategies for novel directions, applications, and techniques.

Eye Anatomy

This section reviews important anatomic structures of the human eye, which builds a foundation for the discussion of computational geometric eye modeling in the following section.

The human eye

The eye is the organ that provides the optics and photo reception for the visual system—and the anatomy follows its function in this physiological process. An outer

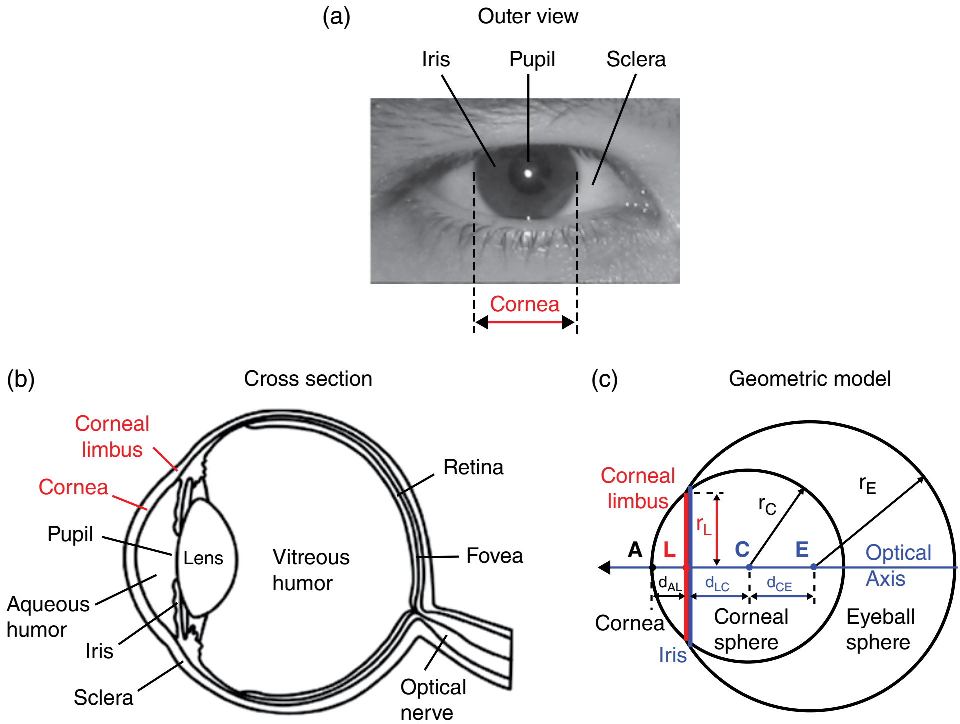


Figure 21.2 Eye model. (a) Outer view and (b) cross section of the human eye. (c) Approximation by a spherical eye model with static parameter values.

view of the eye in Figure 21.2a shows the color-textured iris and the pupil in its center as the most distinctive components. The iris is surrounded by the white sclera, a dense and opaque fibrous tissue that mainly has a protective function.

Eyeball

A cross section of the eyeball in Figure 21.2b reveals that its main part is located behind the skin and components visible from the outside. Geometrically, the eyeball is not a plain sphere; its outer layer can be subdivided into two approximately spherical segments with different radii and separated centers of curvature: the anterior corneal and the posterior scleral segment (Figure 21.2c). The smaller anterior segment covers about one-sixth of the eye, and contains the components in front of the vitreous humor, including cornea, aqueous humor, iris, pupil, and lens. It has a radius of curvature r_C of approximately 8 mm. The posterior segment covers the remaining five sixths with a radius of curvature r_E of approximately 12 mm. Both centers of curvature are separated by a distance d_{CE} of approximately 5 mm. The eyeball is not symmetric; its diameters are approximately 23.5 mm horizontal (d_H), 23 mm vertical (d_V), and 24 mm anteroposterior (d_{AP}) (distance between the anterior pole at the apex of the cornea and the posterior pole at the retina). See Table 21.1 for an overview of parameter values from different sources.

Table 21.1 Eye parameter variation (mm).

	Eyeball			*Posterior*	*Anterior*			
	d_{AL}	d_{LC}	d_{CE}	r_E	r_C	r_L	r_{LH}	r_{LV}
(a) Books on eye anatomy								
Snell and Lemp (1997)	—	—	—	12.00	7.70	5.575[a]	5.85	5.30
Crick and Khaw (2003)	—	—	—	—	—	5.75[a]	6.00	5.50
Kaufman and Alm (2003)	—	—	—	—	7.80	6.075[a]	6.30	5.85
Remington (2004)	—	—	5.70	12.00	7.80	5.75[a]	6.00	5.50
Khurana (2007)	—	—	—	12.00	7.80	5.675[a]	5.85	5.50
(b) Schematic eye models								
Gullstrand (1909) No. 1	3.60	—	—	—	7.80	—	—	—
Gullstrand (1909) No. 2	3.70	—	—	—	7.70	—	—	—
Le Grand and El Hage (1980) 1945	3.60	—	—	12.30	7.80	—	—	—
Lotmar (1971)	3.60	—	—	12.30	7.80	—	—	—
Kooijman (1983)	3.55	—	—	—	7.80	—	—	—
Liou and Brennan (1997)	3.66	—	—	—	7.77	—	—	—
Escudero-Sanz and Navarro (1999)	3.60	—	—	12.00	7.72	—	—	—
(c) Work on eye modeling and applications								
Lefohn et al. (2003)	2.50	5.25	4.70	11.50	7.80	5.80	—	—
Morimoto and Mimica (2005)	3.53[a]	4.17	—	—	7.70	6.47[a]	—	—
Hua et al. (2006)	—	—	—	12.50	7.80	5.50	—	—
Nishino and Nayar (2006, 2004)	2.18	—	—	—	7.80	5.50	—	—
Li et al. (2007)	3.05	4.75[a]	5.70[a]	12.50[a]	7.80	6.19[a]	—	—
This chapter	2.27[a]	5.53[a]	5.70	—	7.80	5.50	—	—

Notes: [a] Calculated from given values.

The eye has several axes—the optical and visual axes are the two major ones. The optical axis is usually defined as the line joining the centers of curvature of the refractive surfaces, connecting corneal apex **A**, limbus center **L**, corneal center **C** and eyeball center **E**. The visual axis describes the gaze direction of the eye. It is defined as the line joining the fovea and the object being viewed, which slightly differs from the optical axis. Both axes intersect at the nodal point of the eye, where the image of the gazed object becomes reversed and inverted. For a typical adult, the deviation of the visual axis is 4°–5° nasal and 1.5° superior to the optical axis with a standard deviation of 3° (Hansen & Ji, 2010).

Cornea

The transparent cornea is the outer layer of the eye that covers the iris and dissolves into the sclera at the corneal limbus. Beside having a protective function, the cornea plays the main role for the eye as an optical system to focus images on the retina. Its transparency and optical clarity stem from three factors (Kaufman & Alm, 2003; Crick & Khaw, 2003): (a) the uniform size and arrangement of submicroscopic collagen fibrils, (b) the absence of blood vessels (avascularity), and (c) the relative state of dehydration. The internal pressure of the eye is higher than that of the atmosphere, which maintains the corneal shape and produces a smooth external surface.

Table 21.2 Anatomic zones of the cornea.

Corneal zone	*Approximate diameter (mm)*	*Characteristics*
Central optical	0–4	most spherical, symmetric; overlies pupil
Paracentral (mid)	4–7	generally spherical, but progressive flattening
Peripheral	7–11	greatest flattening and asphericity
Limbal	11–12	cornea steepens before joining the sclera

Source: Adapted from Snell and Lemp (1997).

In addition, the surface is coated with a thin film of tear fluid that ensures its smoothness and helps to nourish the cornea. As a result, the surface shows mirrorlike reflection characteristics.

Although the corneal surface is approximately a sphere, it has only spherical curvature near the apex and generally flattens towards the periphery. The cornea is subdivided into four anatomic zones with an increasing radius from the optical axis (Table 21.2). The eyeball is usually not rotationally symmetric around the optical axis but slightly flat in the vertical direction. This leads to a toricity in the corneal surface with a higher vertical curvature. Considerable individual variation occurs in eye surface curvature, component separation, and axial length. The typical cornea approximates to an ellipsoid, with an apex radius of curvature r_C of approximately 7.8 mm. Asphericity values for individual eyes are widely distributed and can include some cases where the cornea steepens rather than flattens towards the periphery.

Corneal limbus

The area where the transparent cornea dissolves into the opaque sclera is called the corneal or corneoscleral limbus. It is a band, approximately 1.5–2.0 mm wide, which surrounds the periphery of the cornea (Snell & Lemp, 1997). The radius of curvature immediately changes at this intersection, creating a shallow groove with a shape discontinuity on the outer surface of the eye. Refer to Table 21.1 for an overview of common values for horizontal radius r_{LH}, vertical radius r_{LV}, and mean radius r_L of the limbus.

Histologically, the limbus contains the transition from the regular lamellar structure of collagen fibrils of the cornea to the irregular and random organization of collagen bundles in the sclera. The layers of corneal tissue either merge into scleral tissue or terminate at different landmarks. The limbal area further contains blood vessels and lymphatic channels. This leads to a smooth and nonuniform transition.

Iris

The iris is a thin, pigmented, circular structure located directly in front of the lens. Its mean radius r_I is 6 mm. The outer structures of the iris extend behind the limbus and the beginnings of the sclera. The area visible on the outside is delimited by the transparent corneal tissue that inhomogeneously dissolves at the limbus.

Iris color for normal eyes ranges from light blue to dark brown, depending on the arrangement and density of the connective tissue and pigment. The color may vary between both eyes of the same person and different parts of the same iris (Snell & Lemp, 1997). The surface of a heavily pigmented brown iris appears smooth and velvety, whereas the surface of a lightly colored gray, blue, or green iris appears rough and uneven.

Pupil

The iris forms the diaphragm of the optical system with a central circular aperture, the pupil. The size of the pupil controls retinal illumination with a diameter varying between 1 and 8 mm. In about 25% of individuals it slightly differs in size (Snell & Lemp, 1997). The image of the pupil seen on the outside is a virtual image corresponding to the entrance pupil that is forward to and slightly larger than the real pupil (Atchison & Smith, 2000). Compared to the smooth appearance of the iris boundary, the circular pupillary margin is a rather sharp edge. The pupil appears black, because most of the entering light is absorbed by the tissues of the inner eye. The pupil can appear red in an image, where the eye is photographed with bright flash illumination under low-intensity ambient light. This so-called *red-eye effect* (van de Kraats & van Norren, 2008) is caused by the large amount of light, reflected from the back of the eyeball in the direction of the camera, when the flash is located near the lens (Figure 21.3a).

Eye Model

This section surveys the construction of computational eye models in the context of different applications. Modeling the properties of the eye is required to accomplish the tasks related to corneal imaging.

Parametric model

Knowledge of the shape and parameter distribution of the human eye allows for the construction of parametric eye models. Several so-called *schematic eye models* with different levels of sophistication have been developed over the last 150 years, motivated by the aim to describe the imaging characteristics and performance of the eye as an optical system (Atchison & Smith, 2000; Bakaraju, Ehrmann, Papas, & Ho, 2008; Gullstrand, 1909; Kooijman, 1983; Le Grand & El Hage, 1980; Liou & Brennan, 1997; Lotmar, 1971). These models represent an average eye, created from average population measurements. They are the foundation for most works in optical eye analysis and modeling.

For applications related to corneal imaging, it is usually sufficient to model the outer (visible) surface of the eye, which does not require dealing with refractive surfaces of the inner eye. The outer surface of the eye is modeled as a surface of revolution around the optical axis, with two intersecting surfaces for the eyeball and the cornea. Due to simple geometry and computation, the most common model uses a spherical surface. Accuracy can be increased by modeling asphericity using general quadric surfaces of revolution, such as spheroid (Baker, 1943), ellipsoid (Nakazawa et al., 2016; Nishino & Nayar, 2006; Beymer & Flickner, 2003), general conicoid

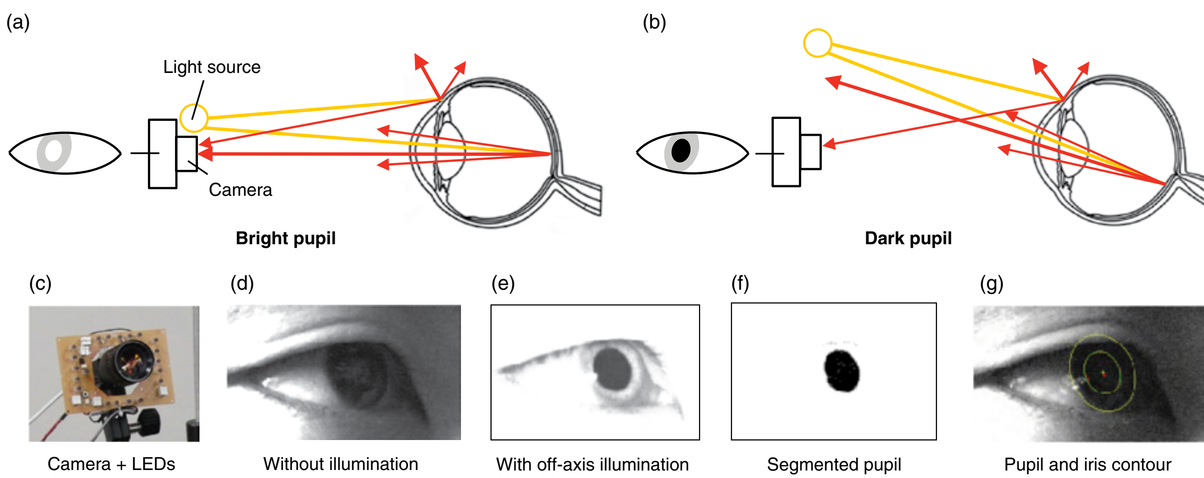


Figure 21.3 Pupil segmentation and iris contour fitting. (a) Bright-pupil effect from on-axis illumination, also known as red-eye effect in flash photography. (b) Dark-pupil effect from off-axis illumination. (c) Camera with off-axis IR LEDs. (d) Eye image without LED illumination. (e) Off-axis illumination creates a dark pupil. (f) Segmented pupil from difference image. (g) Segmented pupil contour, and fitted ellipses to pupil and iris contours.

(Atchison & Smith, 2000), or higher-order surfaces (Gasparini & Caglioti, 2011; Nagamatsu, Iwamoto, Kamahara, Tanaka, & Yamamoto, 2010).

Spherical model In the spherical case, the eye is represented as two overlapping spheres with different radii and separated centers of curvature, **C** and **E** (Figure 21.2c) (Gatinel et al., 2011). For reference, the applied parameter values are listed in Table 21.1. The cornea is modeled as a spherical cap, with a radius of curvature r_C of 7.8 mm (Kaufman & Alm, 2003), which is cut off from the corneal sphere by the limbus plane. The visible part of the iris is assumed to be equal to the circular limbus, with a mean radius r_L of 5.5 mm (Nishino & Nayar, 2006). The displacement d_{LC} between the centers of the limbal circle and the corneal sphere are obtained from the given parameters as in

$$\begin{aligned} d_{\mathrm{LC}} &= \sqrt{r_{\mathrm{C}}^2 - r_{\mathrm{L}}^2} \\ &\approx 5.53\,\mathrm{mm}. \end{aligned} \tag{21.1}$$

The height of the cornea, defined as the distance d_{AL} between the corneal apex **A** and the center of the circular limbus **L**, is obtained as in

$$\begin{aligned} d_{\mathrm{AL}} &= r_{\mathrm{C}} - d_{\mathrm{LC}} \\ &\approx 2.27\,\mathrm{mm}. \end{aligned} \tag{21.2}$$

All eye movements are described as rotations around the geometric center of the eye **E**, located at a distance d_{CE} of approximately 5.7 mm posterior to the center of the corneal sphere, where the limited set of anatomically possible eye poses is described by Donder's and Listing's laws (Tweed & Vilis, 1990). Corneal reflection analysis, however, does not usually require modeling the surface of the eyeball and eye movements.

The centers of curvature of cornea and eyeball, **C** and **E**, and the centers of the optical components corneal apex and pupil, **A** and **L**, all lie on the optical axis of the eye. The true gaze direction, the visual axis of the eye, is described by the fovea and the nodal point of the eye. The position of the nodal point changes whenever the user focuses at a different distance; however, it is generally assumed to coincide with pupil center **L**. The static offset between optical and visual axis is approximately 5°, where the two degrees of freedom (DOF) are described as separate horizontal and vertical offset angles, β and α, or a joint orientation and offset angle κ, from the optical axis.

The model is not perfect, as the user-dependent parameters are assumed to be static, and modeling the cornea and eyeball as a spherical surface is not anatomically exact. Nonetheless, we may use it as a simple but effective approximation that has been successfully applied in several studies (Johnson & Farid, 2007; Nakazawa & Nitschke, 2012; Nitschke et al., 2011a; Nitschke & Nakazawa, 2012; Plopski, Kiyokawa, Takemura, & Nitschke, 2014; Schnieders et al., 2010; Takemura et al., 2014b; Tsumura et al., 2003).

Individual parameter estimation The accuracy of eye analysis and modeling applications can be increased by combining the parametric eye model with individually measured personal parameter values. In the domain of eye-gaze tracking (EGT), the measurement is commonly done by a one-time interactive calibration procedure, asking the user to perform a task like gazing at markers and moving eyes. Some approaches even automatically estimate the parameters during runtime, removing the need for tedious manual interaction and expert knowledge.

Eye-feature-tracking based passive eye-pose estimation methods use simple one-sphere eye models that do not account for the corneal surface. Wu, Kitagawa, Wada, Kato, and Chen (2007) describe a two-eye eye-model tracking with a simple calibration strategy to estimate the personal parameters of interpupil distance, eyeball radius r_E and iris radius r_L. In the first frame, four eye corners of both eyes are manually selected. In subsequent frames, the personal parameters are estimated through particle-filter tracking and known eye gaze, where the person is assumed to look straight forward. Reale, Canavan, Yin, Hu, and Hung (2011) describe a remote point-of-gaze (PoG) tracker with an interactive calibration procedure, where the user gazes at the camera center and two known 3D points on a screen. The system estimates the iris radius r_L, the distance between eyeball center and iris d_{LE}, and the visual axis offset. Tsukada, Shino, Devyer, and Kanade (2011) and Tsukada and Kanade (2012) introduce a wearable eye-gaze tracker with an automatic calibration procedure to determine eyeball radius r_E, iris radius r_L and the location of the static eyeball center **E**. The parameters are iteratively estimated by gradually increasing eye model constraints.

Pupil-center–corneal-reflection (PCCR) based active-light eye pose estimation methods use two-sphere eye models that account for the corneal surface. Villanueva and Cabeza (2008) perform a geometric evaluation for personal calibration in the active light case. They show that two glints, from point light sources, at a known location with respect to the eye-tracking camera, and a single calibration point, are the minimum requirement to calibrate the corneal radius r_C, the cornea–iris distance d_{LC} and the major horizontal optical-visual axis offset angle β. Further one-point calibration methods have been proposed for single-camera (Nakazawa & Nitschke, 2012; Ohno, 2006; Villanueva & Cabeza, 2008) and stereo-camera setups (Guestrin & Eizenman, 2006, 2011; Nagamatsu, Kamahara, Iko, & Tanaka, 2008).

Plopski (2016) and Plopski, Nitschke, Kiyokawa, Schmalstieg, and Takemura (2015b) propose a hybrid eye-pose estimation approach, improving passive eye tracking with reflections from a known scene model (as opposed to IR LEDs in active-light methods). This allows improvement of the accuracy and robustness of passive tracking, and allows estimation of the personal parameters iris size r_L and cornea–iris distance d_{LC}.

Detailed model

It is often sufficient to apply parametric eye models to simplify implementation and computation requirements. However, application scenarios exist that require more detailed and accurate eye modeling, such as eye surgical and optical refractive correction in medicine, and anatomically and physically realistic eye modeling in computer graphics and vision. The following discusses such works from different fields.

Complete eye As accurate and realistic eye modeling is important to various fields, there exists a large body of previous work. Ruhland et al. (2014) provide a recent survey of eye and gaze animation for virtual agents and artificial systems. The work comprises a broad overview from low-level eye physiology, geometry and appearance, to high-level gaze behavior.

Medicine. In medicine, it is important to reconstruct accurate individual eye models regarding anatomy, geometry, optics, appearance, and dynamics. Sagar, Bullivant, Mallinson, and Hunter (1994) develop an anatomically detailed model of the eye and surrounding face for surgical simulation in a virtual environment. The model visually and mechanically simulates the features of the human eye by combining realistic computer graphics with finite element analysis. The accurate shape of the cornea is acquired using laser-based confocal microscopic imaging through a microsurgical robotic system. Einighammer, Oltrup, Bende, and Jean (2009) describe the reconstruction of an individual geometric and optical eye model. The cornea is represented as a spline-interpolation from the measured anterior corneal topography. Then, the geometry of the lens is calculated through wavefront aberration optimization. Priamikov and Triesch (2014) introduce a platform for biomechanical simulation of eye movements. Based on measurements of the eye, they develop a biomechanical model of the human extraocular eye muscles and a visualization of the eyeball.

Computer graphics. In computer graphics, biophysical and photorealistic modeling, rendering, and animation of eyes has a wide range of applications, such as virtual reality simulation and entertainment. Jimenez, Jarabo, Gutierrez, Danvoye, and Pahlen (2012) describe GPU-accelerated approaches for photorealistic eye rendering, covering several anatomical and optical properties, such as wetness, eye redness, ambient occlusion, refraction, and reflection. Bérard, Bradley, Nitti, Beeler, and Gross (2014) present a system to acquire and model the spatio-temporal shape and texture of an eye at very high resolution. They describe a complex illumination and capturing hardware system and hybrid reconstruction strategy to handle the different appearance properties of the visible parts of the eye. Starting from a generic eye proxy, the system recovers the sclera, the cornea and the iris, and combines them into a complete eye model. The method is the current state of the art in high-accuracy geometry and appearance eye modeling using standard optical components.

Computer vision. In computer vision, biophysically and photo-realistic modeling and rendering of eyes has a wide range of applications in HCI, human sensing and biomedical imaging. Świrski and Dodgson (2014) generate synthetic ground-truth eye images to evaluate eye tracking algorithms. Based on Holmberg's public domain head model (Holmberg, 2012), they build a complete 3D head model with a spherical eye, which allows mobile and remote eye tracking systems to be simulated. Physically correct rendering allows handling of light emission, reflection, refraction, shadow, depth-of-field blur, camera-shot noise, and IR imaging. Their system uses the open-source software Blender, supporting GPU-accelerated rendering and Python batch scripting. Wood et al. (2015) show the photorealistic rendering of eye images to generate training data for supervised computer vision applications, which removes the problem of time-consuming and potentially unreliable data collection and manual annotation. They reconstruct dynamic eye-region models, with two-sphere eyeball geometry, from professionally acquired head-scan geometry. The models are then used to render close-up eye images for a wide range of shape and appearance variation, regarding individual characteristics (gender, ethnicity, age), and head pose, face and eye motion, and illumination condition. They show that the synthesized training data (SynthesEyes) outperforms state-of-the-art methods for eye-shape registration (to detect anatomical landmarks) and appearance-based gaze estimation. In the context of corneal reflection analysis, Johnson and Farid (2007) propose a physically based spherical model for the cornea and visible structures of the eye. Nitschke (2011) extends the model with aspherics, where eye structures are modeled as ellipsoids and elliptical cross sections. The model is applied in a scripting framework (Pharr & Humphreys, 2004) for rendering synthetic eye images with corneal reflections of environmental illumination. This allows the simulation of the impact of different parameters on corneal reflection modeling, especially where ground-truth measurements are difficult to obtain, as with parameters related to the individual eye.

Corneal shape Corneal reflection analysis requires a known pose and shape for the cornea. Related works commonly apply spherical models with fixed size or personal parameters recovered using multiple cameras or calibration procedures. Only a few approaches support conicoid or general rotational symmetric models that better describe the corneal periphery (Atchison & Smith, 2000; Guillon, Lydon, & Wilson, 1986; Lindsay, Smith, & Atchison, 1998; Ying, Wang, & Shi, 2012). While most of these rely on approximate values (Beymer & Flickner, 2003; Nishino & Nayar, 2006), a few recent works also estimate a personal asphericity parameter for a conicoid (commonly prolate ellipsoid) representation. In the context of eye-gaze tracking (EGT) using an active-light 3D geometric approach, Nagamatsu et al. (2010) extend the common interactive user calibration to additionally estimate the asphericity as the radius of curvature at the corneal apex. Nakazawa et al. (2016) further show that the asphericity can be recovered automatically. They describe a method for estimating the warping function between a scene image and the corresponding corneal reflection of the scene from an eye image (see the section on corneal reflection–scene matching). The accuracy of the image registration depends on the accuracy of the geometric modeling, where the shape of the corneal reflector has a high impact. To increase accuracy, they refine their initial algorithm (Nakazawa, Nitschke, & Nishida, 2016) to model and automatically estimate asphericity together with eye pose and scene orientation in an iterative nonlinear image registration framework. Though, the work targets corneal reflections from natural scenes, the approach could be also employed with artificial illumination in eye-tracking scenarios.

Regarding measurement of the accurate corneal surface topography, there exist several nonintrusive optical techniques in the context of ophthalmology and optometry. Keratometry (ophthalmometry) considers the cornea to be a spherical reflective surface and measures its radius of curvature. The calculation is based on geometric optics, applied to only four sampling points in a small portion of the central cornea. More accurate shape models may be reconstructed through reflections from controlled illumination using the principle of shape from specular reflection (Balzer & Werling, 2010; Ihrke, Kutulakos, Lensch, Magnor, & Heidrich, 2010). Photokeratoscopy, or videokeratography (Halstead, Barsky, Klein, & Mandell, 1996; Mandell, 1996; Swartz, Marten, & Wang, 2007) is a diagnostic technique that applies this principle to reconstruct accurate models of corneal topography (Bogan, Waring III, Ibrahim, Drews, & Curtis, 1990; Gatinel, Malet, Hoang-Xuan, & Azar, 2011) for various medical applications such as refractive surgery, change monitoring, disease diagnosis, and contact lens development. The technique works as follows: A patient is seated in front of a keratographer, a concave device that displays an illuminated pattern (commonly a series of concentric rings or a moving slit light). The pattern is focused on the anterior surface of the patient's cornea and reflected back to a digital camera at the center of the pattern. This allows the shape of the whole cornea to be reconstructed from the distortion of the reflected pattern at several thousand sampling points. The result can be represented in a number of formats, such as an axial, tangential, elevation, or refractive map, to visualize different characteristics of corneal topography.

Iris appearance and dynamics The iris is important for corneal reflection analysis for various reasons. (a) Iris contour tracking is the common technique for passive eye-pose estimation to determine the pose of a 3D eye model relative to an eye camera. (b) Handling the overlay of iris pattern and scene reflections is required for improving the visual quality and geometric modeling of corneal reflections, and removing scene reflection noise in iris biometrics. (c) Corneal reflection analysis captures the environmental map of incident illumination at the eye, which is related to modeling the pupillary light reflex (PLR) and normalizing pupil dynamics for human internal state estimation. In computer graphics, the requirement of photorealistic rendering and animation of the human iris has led to techniques for the accurate modeling of iris shape, pattern, appearance, and dynamics. Lam and Baranoski (2006) introduce the first biophysically based light transport model for the iris, to simulate the light scattering and absorption of the iridial tissue and calculate the spectral radiometric response. Francois, Guatron, Breton, and Bouatouch (2009) propose a method for photorealistic iris/eye modeling and real-time rendering. The model is obtained by image-based reconstruction from an environment map and an image of the eye. The iris morphology and scattering features are recovered by estimating camera pose and accounting for corneal refraction. Pamplona, Oliveira, and Baranoski (2009) introduce a physiologically based model for the pupillary light reflex and an image-based model for iridal pattern deformation. The model for the pupillary light reflex expresses the pupil diameter as a function of environmental illumination, described by a delay-differential equation with parameter values derived from actual measurements. The model for realistic iridal pattern deformation is then expressed as a function of pupil dilation and constriction.

Eye Pose Estimation

This section covers the position and orientation estimation of the eye model relative to the camera, which is equivalent to 3D model-based eye tracking and calibrating the cornea-camera catadioptric imaging system. Eye-pose estimation recovers the gaze direction up to the optical axis. An additional one-time individual calibration with at least a single calibration point is necessary to recover the offset to the true gaze direction or visual axis (Guestrin & Eizenman, 2006; Nakazawa & Nitschke, 2012; Villanueva & Cabeza, 2008). Table 21.3 shows an overview of different methods, mostly related to eye-gaze tracking (EGT). Accuracy is indicated as the error in gaze direction, a common measure that relates to the pose of the eye and, thus, to the position of the cornea.

Eye-pose estimation requires two tasks: *image processing* and *geometric modeling*. Image processing determines if and where an eye occurs in the image, and tracks the detailed location of particular features that can be real anatomic structures or corneal

Table 21.3 Comparison of eye pose/gaze estimation methods.

			Personal	*Gaze*		*Eye pose estimation method*		
Reference	*Cameras*	*Lights*	*calibration (pts)*	*information*	*error (deg.)*	*Eye position*	*Gaze direction*	*Note*
(a) Passive, academic								
Wang and Sung (2001)	1 + 1	—	—	Optical	0.86 ± 0.16	Eye model	Iris contour†[(1)]	‡(3)
Wang and Sung (2002)	1 + 1	—	—	Optical	0.48 ± 0.09	Eye model	Iris contour†[(2)]	‡(4)
Wu et al. (2005)	1	—	—	Optical	7.12 ± 4.65	—	Iris contour†[(1)]	‡(5)
Nishino and Nayar (2006)	1	—	—	Optical	5.95	Eye model	Iris contour†[(3)]	
Yamazoe et al. (2008)	1*	—	§(1)	Optical	9.19 ± 1.48	Image reprojection error from 3D face/eye model		‡(6)
Chen and Ji (2008)	1*	—	9	Visual	3.34	Eye corners	Pupil center	
Schnieders et al. (2010)	1	—	§(2)	Optical	2.28 ± 0.40	Eye model	Iris contour, scene constraints†[(4)]	
Reale et al. (2010)	1	—	4	Visual	1.09 ± 0.67	Calibration + 3D face pose	Iris back projection error to 3D eye model	‡(3)
(b) Active-light, academic								
Ohno et al. (2002)	1	1	4	Visual	0.70 ± 0.13	Single CR, depth from focus	Pupil contour Iris/pupil contour	
Beymer and Flickner (2003)	2 + 2	2	≥ 2	Visual	0.60	Model fitting using multiple CRs + pupil contour		‡(2)
Hennessey et al. (2006)	1	2	4	Optical	0.73 ± 0.13	Multiple CRs	Pupil contour	
Guestrin and Eizenman (2006)	1	2	9	Visual	0.63 ± 0.10	Multiple CRs	Pupil center	
Villanueva and Cabeza (2007)	1	2	1	Visual	1.08 ± 0.23	Multiple CRs	Pupil contour	
Villanueva et al. (2009)	1	2	5	Visual	1.57 ± 0.51	Multiple CRs	Pupil center	
Guestrin and Eizenman (2011)	2	2 + 2	1	Visual	0.50 ± 0.07	Multiple CRs	Pupil center	

(c) Active-light, commercial							
Smart Eye AB, Smart Eye Pro 5.7 (http://smarteye.se)	1	any	Visual	0.50	Multiple CRs	Head model + iris/ pupil contour	
SensoMotoric Instruments (SMI), SMI RED/RED250/RED500 (https://www.smivision.com)	1	2/5/9	Visual	0.40	Multiple CRs	Pupil	
SensoMotoric Instruments (SMI), SMI IVIEW X™ HED (https://www.smivision.com)	1	5	Visual	<0.50–1.00	Multiple CRs	Pupil	‡(1)
SR Research, EyeLink 1000 (http://www.sr-research.com)	1		Visual	0.50	Multiple CRs	Pupil	
SR Research, SR Research EyeLink II (http://www.sr-research.com)	1		Visual	<0.50	Multiple CRs	Pupil center	‡(1)
Tobii Technology AB Remote Eye Gaze Tracking (https://www.tobii.com)	1	≥ 2	Visual	0.50	Multiple CRs	Pupil	
Tobii Technology Tobii Glasses Eye Tracker (https://www.tobii.com)	1	9	Visual		Multiple CRs	Pupil center	‡(1)

Note: Table 21.3 provides a comparison of eye pose / gaze estimation methods. The column "Cameras" indicates the number of cameras per eye, where "+" refers to additional wide field-of-view cameras and "*" to the use of continuous video frames. The column "Lights" indicates the number of IR LED light sources, where "+" refers to additional sources for redundancy. The column "Personal calibration (pts)" indicates the number of points required for calibrating personal parameters such as visual axis offset, eye size, and cornea–pupil distance. There exist automatic approaches that minimize either §(1) the reprojection error from a 3D face / eye model or §(2) the error from geometric scene constraints. The columns "Gaze information" and "Gaze error (deg)" indicate the obtained axis of the eye and the corresponding error (mean, standard deviation, range). Iris-contour based eye pose estimation commonly results in a two-way ambiguity that is resolved using either †(1) parallel gaze directions in two irises, †(2) equal distance between eyeball center and eye corners, †(3) manual resolution, or †(4) the distance between gaze ray–display intersections. Additional information: ‡(1) Head mounted. ‡(2) Stereo head camera + pan-tilt stereo gaze camera. ‡(3) Head camera + pan-tilt gaze camera. ‡(4) Head camera + pan-tilt-zoom gaze camera. ‡(5) In conjunction with a pan-tilt camera system. ‡(6) Low-resolution images, no intrinsic calibration required. Empty cells indicate that no information is available.

reflections (glints). Eye feature tracking methods commonly require high-resolution eye images that are obtained using stationary high-resolution cameras, camera arrays (Chong, Nitschke, Nakazawa, Rozga, & Rehg, 2017), dynamic pan-tilt-zoom (PTZ) cameras (Yoo & Chung, 2005; Reale et al., 2010), or mirrors (Kim, Sked, & Ji, 2004). Image-based eye detection and tracking is a large topic with a broad range of approaches. Refer to Hansen and Ji (2010) for a recent survey. Geometric modeling comprises algorithms that estimate the pose of a 3D geometric eye model from the image information. In the following, we distinguish between *passive methods,* which work on any eye image, and *active-light methods,* which require additional controlled illumination.

Eye-gaze tracking (EGT) Eye pose estimation is closely related to EGT, the problem of tracking the gaze direction (movement of the eye) or the point of gaze (PoG) in the scene (Duchowski, 2007; Hansen & Ji, 2010; Holmqvist et al., 2011; Young & Sheena, 1975). Eye pose estimation comprises the first part of geometric EGT, as it recovers the eye pose and gaze direction. The PoG is then obtained by either intersecting the gaze ray (optical axis or visual axis, considering the personal offset) with a 3D scene model or applying a calibrated mapping from the eye pose to the scene. For more details, see the section on 3D point-of-gaze tracking.

Active-light methods

Active-light methods are developed for accurate automatic eye-gaze tracking (EGT) and require a complex hardware system with calibrated light sources and eye parameters.[2] The pupil-center–corneal-reflections (PCCR) technique is largely covered in research (Guestrin & Eizenman, 2006; Ohno et al., 2002; Shih, Wu, & Liu, 2000; Villanueva & Cabeza, 2007; Villanueva et al., 2009), and the method of choice in commercial systems (Tobii Technology, Remote Eye Gaze Tracking, https://www.tobii.com; SensoMotoric Instruments (SMI), Gaze & Eye Tracking Systems, https://www.smivision.com; Pupil Labs, Open source eye tracking, https://pupil-labs.com; SR Research, http://www.sr-research.com). The technique involves a two-step approach, first estimating the position of the cornea from light reflections of multiple light sources at known locations, commonly in the form of IR LEDs (Guestrin & Eizenman, 2006; Villanueva & Cabeza, 2007).[3] Using multiple cameras also enables recovering individual anatomic parameters. In a next step, the orientation of the eye model is obtained by detecting a second point on the optical axis, commonly using the center or contour of the pupil. Pupil segmentation is often realized using active IR illumination to exploit the bright- (red-eye) and dark-pupil effects (Agustin, Villanueva, & Cabeza, 2006; Ebisawa, 1998; Morimoto, Koons, Amir, & Flickner, 2000) (Figure 21.3). The strength of either effect depends on different factors, such as the opening of the iris, and the age and ethnicity of the subject. Only a few methods exist that operate under visible light (Vezhnevets & Degtiareva, 2003; Yamazoe et al., 2008) because of the reduced contrast of the pupil contour (Grabowski, Sankowski, Zubert, & Napieralska, 2006). Instead of using segmentation, the pupil contour can be obtained through a radial search from a starting point within the pupil. A popular iterative algorithm for this purpose is the *Starburst* algorithm (Li, Winfield, & Parkhurst, 2005).

[2] Another advantage of using corneal reflections is the possibility to recover an individual aspheric model of the cornea to further increase accuracy (Nagamatsu et al., 2010).

[3] While current methods apply artificial light sources, future methods may directly exploit scene illumination and structure (Plopski et al., 2015b).

Passive methods

Due to their reduced hardware and calibration requirements, passive methods are often applied in low-cost nonprofessional solutions for eye-gaze tracking (EGT) and other applications. They are especially important for corneal imaging techniques as they (a) work on natural eye images, allowing for corneal reflections of light from the surrounding scene, and (b) do not require complex calibrated hardware, allowing for existing imagery to be processed from unknown setups. While these properties make them inherently less robust, compensation strategies have been developed to handle noisy eye feature tracking and commonly unknown individual parameters. Using multiple cameras, for example, enables the recovery of individual anatomic parameters and further constrains the gaze direction (Beymer & Flickner, 2003; Kohlbecher et al., 2008; Villanueva & Cabeza, 2007). Another strategy is to employ combined 2D–3D eye-model tracking rather than pure 2D image-based tracking.

Passive methods commonly estimate the pose of the eye from the contour of the iris (Nishino & Nayar, 2006; Nitschke et al., 2011a; Schnieders et al., 2010; Wang & Sung, 2001; Wu et al., 2005), possibly in combination with other features, such as eye corners (Wang & Sung, 2002; Wu et al., 2007) and lids (Wu et al., 2007). In the following, we first review 2D eye feature tracking approaches and the estimation of the 3D eye pose from the observed image features. We then move towards more robust combined 2D–3D eye-model tracking approaches that estimate the 3D eye pose from the correlation of the observed image features and the reprojected image features according to the 3D eye model pose.

2D eye-feature tracking Although individual variation exists, the iris is approximately circular. Under perspective projection, a circle with an arbitrary 3D pose maps to a general ellipse in an image (Hartley & Zisserman, 2003; Semple & Kneebone, 1952). Methods for iris segmentation are proposed in the context of eye tracking (Hansen & Ji, 2010), iris recognition (Bowyer, Hollingsworth, & Flynn, 2008; Matey, Broussard, & Kennell, 2010), and medical imaging (Barry, Pongs, & Hillen, 1997; Iskander, 2006; Iskander, Collins, Mioschek, & Trunk, 2004). The methods either directly fit a shape model to continuous image features, such as intensity gradients and edge distances (Arvacheh & Tizhoosh, 2006; Nishino & Nayar, 2006), or first segment a particular feature and subsequently fit a shape model using least squares (Fitzgibbon, Pilu, & Fisher, 1999; Halir & Flusser, 1998). Regarding the latter, common strategies apply a vertical edge operator to an upright face image (Colombo, Comanducci, & Bimbo, 2007; Wang & Sung, 2002) or perform contour detection along radial directions starting at the approximate center (Barry et al., 1997; Iskander, 2006; Iskander et al., 2004). Adaptions of the Starburst algorithm from pupil contour detection also exist that do not require the starting point to be the approximate center of the iris (Reale et al., 2010; Ryan, Woodard, Duchowski, & Birchfield, 2008).

3D eye-pose estimation from 2D features The pose of the 3D eye model is defined by the pose of the circular limbus, described by the center point $\mathbf{L} = (L_x, L_y, L_z)^{\mathrm{T}}$ and the normal vector $\mathbf{g} = (g_x, g_y, g_z)^{\mathrm{T}}$. Eye-pose estimation aims in recovering these values. As the corneal limbus coincides with the contour of the visible iris, its pose is obtained from the elliptical contour of the imaged iris. There exists a large body of works on closed-form solutions to the monocular reconstruction of circles with application to camera and object pose estimation (Kanatani & Liu, 1993; Safaee-Rad, Tchoukanov, Smith, & Benhabib, 1992; Zheng, Ma, & Liu, 2008). Multiple parallel or coplanar circles additionally allow for increased stability (Gurdjos, Sturm, & Wu, 2006) and estimating camera parameters (Chen, Wu, & Wada, 2004). These works are the basis for several limbus pose estimation algorithms (Nitschke, 2011; Schnieders et al., 2010; Wang & Sung, 2001; Wu et al., 2005). A simpler method assuming weak perspective projection may be applied when the distance between the eye and the camera is much larger than the scale of the eye, as in common photography of people and faces (Nishino & Nayar, 2006; Nitschke et al., 2011a). Using a single camera, the 3D pose of a circle is estimated up to a two-way ambiguity that is resolved through further knowledge. In the context of eye-pose estimation, such constraints are obtained for a single eye image from anthropometric properties (Wang & Sung, 2002) or by assuming a gaze direction towards the camera (Johnson & Farid, 2007), for a singe face image by assuming parallel irises when focusing far-away objects (Wang & Sung, 2001; Wu et al., 2005) or an intersection of the gaze rays at a known display plane (Schnieders et al., 2010), and for multiple face images using geometric scene constraints (Nitschke et al., 2011a). Figure 21.4 shows results of limbus-based eye pose estimation for different subjects and reflected scenes.

Having recovered the pose of the limbus circle, the position of the cornea is then modeled as a spherical surface, described by the radius r_C and center $\mathbf{C}$, located at a distance d_{LC} from the limbus center $\mathbf{L}$, and obtained as in

$$\mathbf{C} = \mathbf{L} - d_{LC}\mathbf{g}. \tag{21.3}$$

Joint 2D–3D eye-model tracking Unfortunately, 2D iris contour tracking commonly fails under conditions of practice, due to occlusion from eye lashes and eye lids (especially towards the eye corners), and the low-contrast gradual transition between iris and sclera compared to the high contrast edges commonly occuring in corneal reflections. Furthermore, there is the intrinsic anatomic limitation that even an accurately detected iris contour does not necessarily coincide with the corneal limbus. To account for the shortcomings of pure 2D feature-based approaches, the problem can be further constrained to require the tracked features to be consistent with the motion of an underlying 3D eye model. In the following, we will lists the most important developments for increasing robustness by combining the 2D feature tracking and 3D pose estimation tasks.

Figure 21.4 Limbus-based eye pose estimation, with iris contour (ellipse), iris center (dot mark, at iris ellipse center), corneal center (dot mark, displaced from iris ellipse center) and gaze direction (line, starting at iris ellipse center).

Yamazoe et al. (2008) propose a framework for remote long-distance 3D head–eye model tracking. The eye tracking comprises calibration-free estimation of the individual eyeball and iris radius, and framewise eye-pose estimation. The parameters are estimated by minimizing the reprojection error between segmented eye images and eye model projections assuming continuous head motion. The method achieves robust gaze direction estimation, with a reasonable mean error of 6° at an eye region size of only 30×15 pixel. Reale, Canavan, Yin, Hu, and Hung (2011) propose an improved framework for remote 3D head–eye model tracking, with the differences of inversely mapping from image to 3D model, performing a one-time calibration of individual parameters, and avoiding unreliable image segmentation. The calibration process requires the user to look at the camera and two points with known 3D location, and estimates the eyeball center relative to the head model, the eyeball radius, the iris radius, and the visual axis offset. These are then applied for framewise gaze direction estimation. First, the 2D iris center and contour are detected. After obtaining the eyeball position relative to the head pose, the eyeball rotation is estimated by minimizing the reprojection error between a rendered eye model template, directed towards the camera, and a rotated-eyeball model with projective texture-mapped eye image. The estimated eyeball pose allows the 2D iris center and contour to be calculated. Then, the gaze direction is obtained by projecting the iris contour to the eyeball sphere, determining the optical axis and the 3D iris center, and compensating for the visual axis offset. Takemura et al. (2014a, b) describe a simple and calibration-free approach for gaze direction estimation with a wearable eye camera. Assuming static eye parameters and a constant eyeball position, the framewise eyeball rotation is estimated by minimizing the reprojection error of the 3D eye model iris region in the inverted binarized eye image. El Hafi, Ding, Takamatsu, and Ogasawara (2016) further improve the performance of the optimization problem using a greedy algorithm.

Tsukada et al. (2011) explain a real-time 3D eye model based iris contour tracking for a wearable gaze tracking system. The approach reduces the problem from 5-DOF ellipse tracking to 2-DOF position estimation under weak-perspective projection of a 3D eye model. The method assumes precalibrated personal parameters comprising eyeball position, and eyeball and iris radius, and a static relation between eye camera and head. This enables a database of iris contour projections to be built and discretely sampled across horizontal and vertical eyeball rotations. The iris contour tracking consists of a three-step process, comprising appearance-based iris detection and edge point extraction, initial eye pose selection through database lookup, and an iterative rough-to-fine iris contour estimation by eye-pose adjustment according to the gradient image. Tsukada and Kanade (2012) improve their method by adding an automatic iterative eye model parameter estimation. For a series of training images under varying eye pose, an initial iris contour is estimated through the described appearance-based iris detection and contour fitting. Then the 3D eye-model parameters are iteratively estimated in a coarse-to-fine strategy, by refining the iris contour estimates and gradually adding eye model constraints regarding eye structure and eye-camera relation.

Pires, Devyver, Tsukada, and Kanade (2013) explain a robust real-time 3D eye-model-based iris contour tracking for a wearable gaze tracking system. The approach reduces the problem from 5-DOF ellipse tracking in the eye image to 3-DOF circle tracking in an unwrapped image of the spherical eye surface between the two eye corners. The method requires a one-time eye model calibration, using the eye corners in a single frame to determine the static position, torsional rotation and radius of the eyeball up to scale. The parameters allow mapping the eye image into an unwrapped spherical eye image, where the iris is circular. The iris is then detected using a three-step robust circle-fitting procedure, comprising an initial detection through a circular Hough transform, a Starburst-like radial search from the iris center, and a robust circle fitting to the detected edge points. Pires, Hwangbo, Devyver, and Kanade (2013) improve the method under conditions of practice by allowing for dynamic movement of the eye tracking camera relative to the head. The simple extension involves continuous eye position calibration through eye corner tracking using template matching.

Wu et al. (2007) describe a sophisticated 3D eye model tracking for gaze and blinking estimation. The method tracks the iris and eyelid contours of both eyes, using a particle filter (Isard & Blake, 1998) with a comprehensive eye model comprising both eyeballs, iris contours, and upper and lower eyelids. Assuming synchronized movements of both eyes allows increased robustness and efficiency, achieving real-time processing. The method requires a simple one-time personal parameter calibration to estimate interpupil distance, eyeball radius, and iris radius. In the first frame, therefore, the four eye corners are manually selected. In subsequent frames, the parameters are estimated through tracking a moving user who looks into the camera.

Based on this, Nakazawa, Nitschke, and Nishida (2015) describe a particle-filter 3D eye model tracking approach for real-time gaze and blinking estimation using a wearable eye camera. The method is calibration free, which means that it assumes constant or precalibrated personal eye parameters and does not require internal camera parameters by using weak-perspective projection. The 2D eye model consists of eyelids and iris. Upper and lower eyelids are modeled as deformable 2D Bezier curves, defined by the eyelid corners, initially detected through learned cascade classifiers, and upper and lower control points. The iris is modeled as an elliptical curve segment, delimited by the eyelids. A two-step particle-filter tracking estimates the 2D eyelid contours and the reprojected 3D iris contour that maximize the likelihood regarding the input image. Here, the ill-posed 2D–3D iris circle reconstruction is disambiguated by the known direction to the center of the cornea, which is obtained from the mean reflection of two micro LEDs mounted above and below the camera lens. As the LEDs may be replaced by any passive markers that can be robustly detected within the corneal image, the method itself is considered passive.

Plopski et al. (2015a) and Plopski (2016) introduce a corneal-imaging based 3D eye model tracking to achieve automatic spatial calibration and eye-gaze tracking in optical see-through head-mounted displays (OST-HMDs). The method resembles 3D eye-model-based PCCR, in that it uses a two-step process to estimate the cornea position and the eyeball position or gaze direction. As in PCCR, the cornea position is estimated from 3D–2D correspondences between corneal reflections of known scene locations, with the difference of using the HMD screen, rather than IR light sources. It therefore requires a precalibrated relationship of eye camera and HMD screen, which can be obtained through a catadioptric or scene-camera–marker approach. The eyeball position and gaze direction estimation does not require pupil or iris contour tracking. Instead, it recovers the eyeball position as the center of the rotational orbit of corneal centers, obtained from at least three frames under eye motion and a constant relation between HMD and head. A benefit of the method is that it does not require dedicated eye region detection, which is implicitly obtained through the detection of the screen reflection.

Table 21.4 Comparison of eye pose estimation approaches.

		Active light		*Hybrid*		*Passive*
Requirements	–	Image (Artificial) scene geometry Synchronized illumination	∘	Image Scene geometry	+	Image
Features	+	Eye reflection Scene reflection	∘	Eye Scene reflection	–	Eye
Accuracy	+	High	∘	Medium to high	–	Low

+ advantage | ∘ neutral | – disadvantage

Plopski et al. (2015b) and Plopski (2016) further develop the idea of *hybrid eye-pose estimation* combining passive tracking with scene correspondences from corneal reflections (Table 21.4). They generalize and improve the method by replacing the HMD screen with an arbitrary known 3D scene, reducing the processing to a single frame, and estimating the personal eye parameters iris size and cornea–iris distance. A 3D scene model may be available as it is required for various application scenarios, or can be acquired in advance or at runtime using single-camera structure-from-motion (SfM), multiple-camera stereo or a depth camera (KinectFusion, Newcombe et al., 2011). The relative pose of the eye camera can be continuously estimated through an attached scene camera. As in Plopski et al. (2015a), the method also uses a two-step process, where the corneal position is estimated from 3D–2D scene correspondences. The reprojection of the recovered cornea then constraints iris contour tracking for eye-pose estimation. The improvements increase robustness and accuracy, especially at large gaze angles. For more details, see the section on head-mounted display calibration.

Corneal Reflection Modeling

This section builds on eye modeling and pose estimation to describe the back-projection for the cornea–camera catadioptric imaging system. This includes a reflection model to determine the direction towards a scene point and the triangulation of rays under multiple eye poses to recover the corresponding scene point position. After that, we provide an overview of the forward-projection problem.

Cornea–camera catadioptric imaging system

A camera capturing an image of the eye that exhibits corneal reflections of the environment can be modeled as a catadioptric imaging system and, thus, benefit from an extensive theory and literature coverage (Sturm et al., 2011) (Figure 21.5). Catadioptric systems combine external mirror(s) into the optical path of a camera to achieve particular imaging characteristics. While a perspective camera has a single viewpoint, where all projection rays intersect, catadioptric systems either have a single (Baker & Nayar, 1999; Bruckstein & Richardson, 2000; Nayar & Peri, 1999) or multiple viewpoints (Swaminathan, Grossberg, & Nayar, 2006), referred to as central or noncentral catadioptric systems, respectively. Calibration denotes the task of determining the projection function, which comprises camera parameters, mirror pose

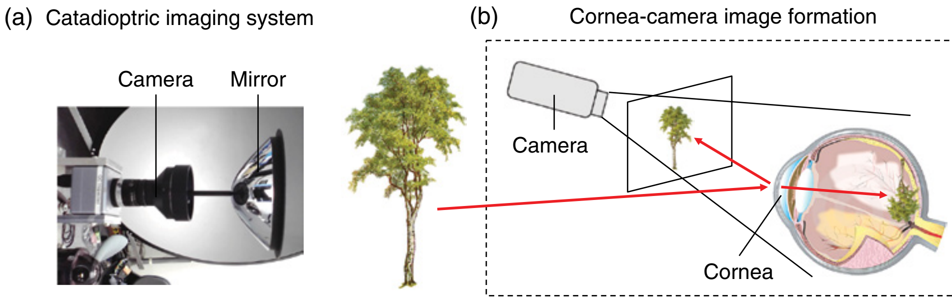


Figure 21.5 Catadioptric imaging. (a) Catadioptric imaging system, where light from the scene reflects at the mirror into the camera. (b) The largest part of light arriving at the cornea refracts and enters the eye. The remaining part reflects back into the environment. When captured by a camera, the compound of cornea and camera acts as a noncentral catadioptric imaging system that requires per-frame calibration through eye pose estimation.

(Lhuillier, 2008; Strelow, Mishler, Koes, & Singh, 2001) and shape (Balzer & Werling, 2010; Caglioti, Taddei, Boracchi, Gasparini, & Giusti, 2007; Ihrke et al., 2010). The cornea, which is shaped similar to an ellipsoid, forms a noncentral catadioptric system that requires per-frame calibration through eye pose estimation. For a specular mirror, pose estimation is achieved from its apparent contour and tracked scene correspondences. For the corneal mirror, this may not be possible due to self-occlusion by the eyeball and superimposed iris texture. Methods, therefore, exploit unique eye features and reflections from known scene points.

Catadioptric stereo system. A catadioptric system with two or more mirrors captures a scene from different viewpoints. This forms a catadioptric stereo system that allows 3D reconstruction even from a single image (Lanman, Crispell, Wachs, & Taubin, 2006a; Nene & Nayar, 1998). The epipolar geometry, describing the relation of two views in stereo vision, also exists for catadioptric stereo systems. In single-viewpoint systems (Geyer & Daniilidis, 2002; Sturm, 2002; Svoboda & Pajdla, 2002), the epipolar plane intersects the surface of the second mirror in a conic section curve that projects to a conic section epipolar curve in the image (Svoboda & Pajdla, 2002). In multiple viewpoint systems (Sturm et al., 2011; Swaminathan et al., 2006), the epipolar curve is commonly obtained using numerical approaches (Würz-Wessel, 2003). Capturing corneal reflections from multiple eye poses creates a noncentral catadioptric stereo system. However, the small size of the cornea and the relatively large distance to the camera allow for central approximation with a single focal point at the center of the limbus (Nishino & Nayar, 2006).

Corneal surface reflection

Assume that light from a scene point **P** reflects at a corneal surface point **S** into the direction of the camera.[4] We want to develop a corneal reflection model to calculate the inverse light path from the image, to determine the direction towards the scene point, located at the unknown position **P** (Figure 21.6). Assuming the surface of the

[4] For this task, it is not necessary to distinguish between an actual light source that radiates light and a scene location that reflects incident light from the environment. Thus, we will use both terms interchangeably.

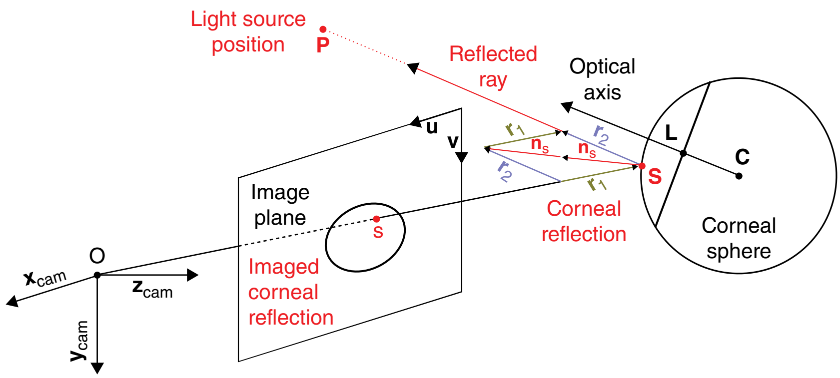


Figure 21.6 Inverse light path towards a point light source. The back-projected light ray from the camera image intersects the corneal surface and reflects into the direction of a light source at an unknown distance from the eye.

cornea to be a perfect mirror, then light from position **P** specularly reflects at surface point **S** into the direction of the camera, where $\mathbf{S} = (S_x, S_y, S_z)^T$ is described as in

$$\mathbf{S}(\phi,\theta) = \mathbf{C} + r_C \begin{bmatrix} \sin\theta\cos\phi \\ \sin\theta\sin\phi \\ \cos\theta \end{bmatrix}, \tag{21.4}$$

with the angles of longitude $\phi \in [0, 2\pi)$ and colatitude $\theta \in [0, \pi]$. An image of the eye captures this specular reflection as a bright patch (glint) or image feature located within the bounds of the visible iris. Let $\mathbf{s} = (s_u, s_v, 1)^T$ denote the subpixel location of the patch centroid or feature in the image, and **S** be modeled as ray $\mathbf{S} = t_1\mathbf{r}_1$ at an unknown distance t_1 from the camera. Here, $\mathbf{r}_1 = \mathrm{K}^{-1}\mathbf{s} / \left\|\mathrm{K}^{-1}\mathbf{s}\right\|$ is the normalized back-projection vector in the direction of **S**, and K the 3×3 camera (projection) matrix that contains the intrinsic camera parameters obtained through calibration. To recover the point of reflection **S**, we calculate the intersection with the corneal sphere by solving the quadratic equation $\left\|\mathbf{S} - \mathbf{C}^2\right\| = r_C^2$ for t_1. Expanding and rearranging leads to

$$t_1^2\mathbf{r}_1^2 - 2t_1(\mathbf{r}_1 \cdot \mathbf{C}) + \mathbf{C}^2 - r_C^2 = 0 \tag{21.5}$$

from which we construct the simplified quadratic formula

$$t_1(\mathbf{r}_1 \cdot \mathbf{C}) \pm \sqrt{(\mathbf{r}_1 \cdot \mathbf{C})^2 - \mathbf{C}^2 + r_C^2}. \tag{21.6}$$

The first intersection at the front side of the cornea is described by the smaller value of t_1. Knowing **S** and the corresponding surface normal $\mathbf{n}_S = (\mathbf{S} - \mathbf{C}) / \left\|\mathbf{S} - \mathbf{C}\right\|$, the normalized direction vector $\mathbf{r}_2$ of the reflection ray is obtained by calculating the specular reflection as in

$$\mathbf{r}_2 = 2(-\mathbf{r}_1 \cdot \mathbf{n}_S)\mathbf{n}_S + \mathbf{r}_1. \tag{21.7}$$

Scene point position **P** then lies on the reflection ray extending from **S**, defined as $\mathbf{P} = \mathbf{S} + t_2\mathbf{r}_2$, at an unknown distance t_2. Registering the reflection rays for the complete iris region at a unit sphere around the cornea creates an environment map of incident illumination that allows for illumination normalization in computer vision tasks, and for extracting the visual information of the scene as a spherical panorama or a virtual perspective camera image (see the sub sections on environmental map reconstruction and local tangent plane projection in the following section).

Scene point position estimation

From a single eye image, we obtain the direction towards a scene location at an unknown distance. To reconstruct its position, we triangulate corresponding light paths from multiple eye images. Therefore, we capture a set of images with varying eye poses for a static scene point. Its unknown position **P** is obtained as the intersection of $N \geq 2$ inverse reflection rays, by estimating the point with minimal distance to the set of rays (Figure 21.7).

Regarding application to corneal imaging, Nishino and Nayar (2006) show that already two eyes, captured in a single face image, enable the reconstruction of simple 3D structure. This idea is further developed to reconstruct the model and pose of a planar computer screen from two or more images using point (Nitschke et al., 2011a) or line features (Schnieders et al., 2010).

The following explains two approaches for reconstructing a 3D scene point from its reflection in two or more images.

Geometric approach for $N = 2$ There exists a simple geometric approach for the triangulation of two rays in 3D. The idea is to compute **P** as the midpoint of the shortest line connecting the two rays

$$\begin{aligned} \mathbf{P}_1 &= \mathbf{S}_1 + t_{21}\mathbf{r}_{21}, \\ \mathbf{P}_2 &= \mathbf{S}_2 + t_{22}\mathbf{r}_{22}. \end{aligned} \tag{21.8}$$

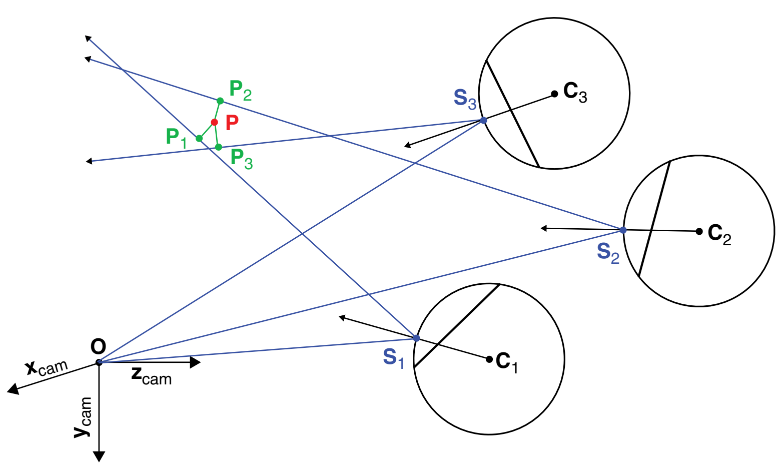


Figure 21.7 Light source position estimation as the intersection of multiple inverse reflection rays. Since the rays generally do not intersect in a single point, we find the least-squares approximation as the point **P** with minimal distance to the set of rays.

From the orthogonality constraint for the shortest connecting line we obtain the two equations

$$\begin{aligned}(\mathbf{P}_1 - \mathbf{P}_2)\cdot \mathbf{r}_{21} &= 0,\\ (\mathbf{P}_1 - \mathbf{P}_2)\cdot \mathbf{r}_{22} &= 0,\end{aligned} \tag{21.9}$$

that are solved for t_{21} and t_{22}. Inserting the ray equations (21.8) into the constraints (21.9) and expanding the dot product leads to

$$\begin{aligned}(\mathbf{S}_1 - \mathbf{S}_2)\cdot \mathbf{r}_{21} + t_{21}(\mathbf{r}_{21}\cdot \mathbf{r}_{21}) - t_{22}(\mathbf{r}_{22}\cdot \mathbf{r}_{21}) &= 0,\\ (\mathbf{S}_1 - \mathbf{S}_2)\cdot \mathbf{r}_{22} + t_{21}(\mathbf{r}_{21}\cdot \mathbf{r}_{22}) - t_{22}(\mathbf{r}_{22}\cdot \mathbf{r}_{22}) &= 0.\end{aligned} \tag{21.10}$$

Solving for t_{21}, backsubstituting, and then solving for t_{22} gives

$$\begin{aligned}t_{21} &= \frac{\big((\mathbf{S}_1 - \mathbf{S}_2)\cdot \mathbf{r}_{22}\big)(\mathbf{r}_{22}\cdot \mathbf{r}_{21}) - \big((\mathbf{S}_1 - \mathbf{S}_2)\cdot \mathbf{r}_{21}\big)(\mathbf{r}_{22}\cdot \mathbf{r}_{22})}{(\mathbf{r}_{21}\cdot \mathbf{r}_{21})(\mathbf{r}_{22}\cdot \mathbf{r}_{22}) - (\mathbf{r}_{22}\cdot \mathbf{r}_{21})(\mathbf{r}_{22}\cdot \mathbf{r}_{21})},\\ t_{21} &= \frac{\big((\mathbf{S}_1 - \mathbf{S}_2)\cdot \mathbf{r}_{22}\big) + t_{21}(\mathbf{r}_{22}\cdot \mathbf{r}_{21})}{(\mathbf{r}_{22}\cdot \mathbf{r}_{22})}.\end{aligned} \tag{21.11}$$

Finally, the searched point with minimal distance to both rays is obtained as

$$\mathbf{P} = \mathbf{P}_1 + \frac{\mathbf{P}_2 - \mathbf{P}_1}{2}. \tag{21.12}$$

Note that when the denominator t_{21} becomes zero, both rays are parallel and do not intersect. Practically, this case does not occur because different eye poses result in different reflection directions. Nevertheless, it is beneficial to increase the baseline (distance) between the cornea positions as this increases the denominator and, thus, the numerical stability.

Algebraic approach for $N \geq 2$ In the general case, $\mathbf{P}$ can be obtained using matrix algebra as follows: At frame l, the distance between $\mathbf{P}$ and the nearest point on the ray $\mathbf{P}_l = \mathbf{S}_l + t_{2l}\mathbf{r}_{2l}$ is defined as

$$\|\mathbf{P}_l - \mathbf{P}\| = \frac{\|\mathbf{r}_{2l} \times (\mathbf{S}_l - \mathbf{P})\|}{\|\mathbf{r}_{2l}\|}. \tag{21.13}$$

Knowing $\|\mathbf{r}_{2l}\| = 1$ and rearranging leads to

$$\|\mathbf{P}_l - \mathbf{P}\| = \|[\mathbf{r}_{2l}]_\times \mathbf{P} - \mathbf{r}_{2l} \times \mathbf{S}_l\|, \tag{21.14}$$

where $[\mathbf{r}_{2l}]_\times$ represents vector $\mathbf{r}_{2l}$ as a skew-symmetric matrix, given by

$$[\mathbf{r}_{2l}]_\times = \begin{bmatrix} 0 & -z_{\mathbf{r}_{2l}} & y_{\mathbf{r}_{2l}} \\ z_{\mathbf{r}_{2l}} & 0 & -x_{\mathbf{r}_{2l}} \\ -y_{\mathbf{r}_{2l}} & x_{\mathbf{r}_{2l}} & 0 \end{bmatrix}, \tag{21.15}$$

which expresses the cross product as a matrix multiplication. To solve for $\mathbf{P}$ we combine the N equations and formulate the problem as a least-squares minimization in the form $\|\mathrm{A}\mathbf{P} - \mathbf{b}\|$. Finally, point **P** is estimated by solving the system, for example, through the pseudo inverse as in

$$\mathbf{P} = \left(\mathrm{A}^{\mathrm{T}}\mathrm{A}\right)^{-1}\mathrm{A}^{\mathrm{T}}\mathbf{b}, \quad \mathrm{A}_{3N\times 3} = \begin{bmatrix} [\mathbf{r}_{21}]_{\times} \\ \vdots \\ [\mathbf{r}_{2N}]_{\times} \end{bmatrix}, \quad \mathbf{b}_{3N\times 1} = \begin{bmatrix} \mathbf{r}_{21} \times \mathbf{S}_1 \\ \vdots \\ \mathbf{r}_{2N} \times \mathbf{S}_N \end{bmatrix}. \tag{21.16}$$

Forward projection

So far, we have covered the back projection of corneal reflections to estimate the direction and position of scene points. Applications in corneal reflection modeling may also require a solution to the inverse problem of forward projection from the scene, for example to calculate the reprojection error or photometric similarity in estimation, registration and bundle-adjustment tasks, which may allow for the combined estimation of eye poses, corneal shape and scene structure. The problem is more difficult because we need to find the point of reflection *without* knowing the direction of the incident light ray. While commonly solved iteratively (Goncalves & Araujo, 2004; Sturm et al., 2011), recent research developed analytic solutions for fast and accurate calculation: Vanderportaele (2006) models the problem for general and quadric-shaped mirrors using polynomials and studies their roots. Agrawal, Taguchi, and Ramalingam (2011) provide a comprehensive theory on noncentral catadioptric projection. The approach first transforms the problem into the plane of reflection; and then applies the two constraints that (a) the solution lies on the intersection curve with the mirror surface, and that (b) the law of reflection requires equal reflection angles and the reflected ray $\mathbf{r}_2$ to pass through **P**. They show that the solution for quadric-shaped mirrors requires solving a sixth-order polynomial equation. For the special case of a spherical mirror, this reduces to a fourth-order equation that can be solved in closed form, which has been shown earlier by Eberly (2008). Nakazawa and Nitschke (2012), Nitschke (2011), and Nitschke et al. (2013a) derive another formulation for a spherical mirror that additionally handles scene locations at an unknown or approximately known distance. The proposed methods enable an approach for 3D point of gaze (PoG) estimation in arbitrary dynamic environments (see the section on 3D point-of-gaze tracking). For further reading on catadioptric projection, we recommend the extensive survey by Sturm et al. (2011).

Performance evaluation

In the context of display pose reconstruction from corneal reflections (see the section on display–camera calibration), several comprehensive experimental studies analyze the impact of parameter variation in corneal reflection modeling using real (Nitschke, 2011; Nitschke et al., 2011a; Nitschke, Nakazawa, & Takemura, 2009a,b) and synthetic (Nitschke, 2011) data. The main factor is the accuracy of surface normals in catadioptric reflection. In line with this, the findings show a large impact of eye-pose estimation and individual eye geometry on the overall accuracy. The results from basic geometric modeling can be considerably improved by optimization, subject to geometric scene constraints using multiple eye poses. Particular achievements from this strategy are a lower reconstruction error with tolerance to noisy measurements and the breaking of an inherent ambiguity in iris-contour-based eye pose estimation.

Regarding the findings, promising strategies to increase accuracy are (a) to apply the PCCR concept from active-light to passive eye pose estimation by using scene constraints, (b) to calibrate / estimate individual eye parameters, and (c) to process only reflections near the corneal apex, where the cornea is most spherical.

Corneal Image Processing

This section builds on the eye and scene modeling to cover practical application-oriented tasks in corneal image processing.

Environment map reconstruction

Taking an image of the eye captures the specular reflection of the scene within the bounds of the visible iris. Using the developed corneal reflection model allows the recovery of the inverse light paths for the complete iris region. Since the distance t_2 between the corneal surface and the scene is usually much larger than the size of the cornea r_C, it is feasible to register each reflection ray into an environment map of incident illumination, located at the corneal center, as in $\mathbf{P}_{EM} = \mathbf{C} + \mathbf{r}_2$ (Figure 21.8).[5] The environment map can be visualized as a spherical panorama comprising the complete scene reflection (Figure 21.9). Furthermore, it encodes the directional information of incident illumination at the cornea, which can be used for a range of vision tasks, including illumination normalization and photometric stereo.

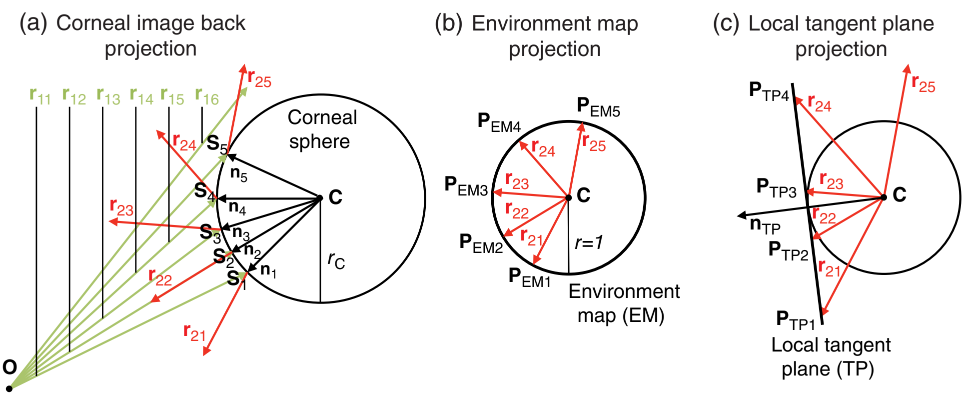


Figure 21.8 Corneal image back projection. (a) Corneal image back projection, comprising the visual information from all pixels within the region of the visible iris. (b) Registration of the back-projected rays into an environment map (EM) of incident illumination at the cornea. This creates a spherical panorama containing the complete scene reflection. (c) Projection of the environment map onto a local tangent plane (TP) at a region of interest. This creates a virtual perspective camera image containing a partial scene reflection.

[5] The maximal distance d_{max} between both rays is r_C, which is easily verified as follows. Since the rays are parallel and $\mathbf{P}$ is located outside the corneal sphere, the distance d can be calculated as the distance between line $\mathbf{C} + t\mathbf{r}_2$ and point $\mathbf{S}$, as in $d = r_C|\mathbf{r}_2 \times \mathbf{n}_S| = r_C \sin(\alpha)$, where $\alpha = \angle(\mathbf{r}_2,\mathbf{n}_S)$. From $\alpha \in [0, \pi/2]$ it then follows that $d \in [0, r_C]$. Furthermore, note that in the context of corneal reflections, α is commonly much smaller than the maximum value.

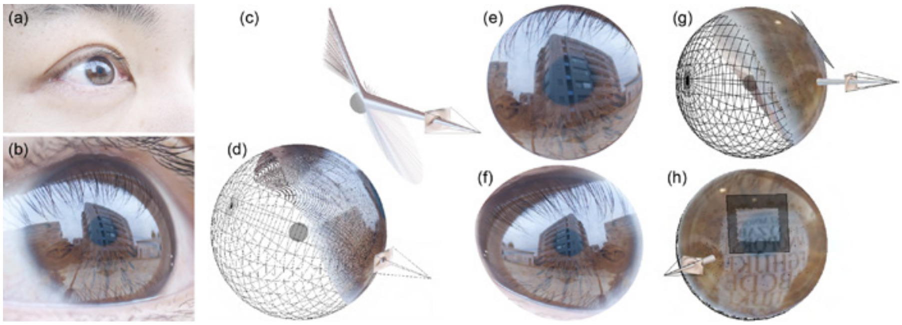


Figure 21.9 Corneal reflection modeling. (a) Eye image (3872×2592 pixel), showing corneal reflections from an outdoor environment containing several buildings. (b) Cropped cornea region (approximately 600×600 pixel). (c) Back projection of limbus pixels, intersecting and reflecting at the corneal surface. (d) Environment map (EM): Back projection of all pixels in the iris region, registered at a sphere around the cornea. (e) EM, outside view towards the cornea. (f) EM, inside view from the cornea. (g),(h) EM and local tangent plane (TP) at region of interest, from corneal reflection of an indoor environment containing a poster with the letters of the alphabet.

Local tangent plane projection

While the environment map (EM) captures the complete scene information in a spherical panorama, the image is inherently distorted. An undistorted partial image can be generated by projecting the environment map into a virtual perspective camera, centered at the environment map sphere and pointing towards a point of interest. The image spans a certain region of interest, modeled as a local tangent plane (TP) to the environment map sphere at the point of interest (Figure 21.8c). Therefore, we set the plane coordinate frame as in

$$\begin{aligned} \mathbf{z}_{TP} &= \mathbf{n}_{TP}, \\ \mathbf{y}_{TP} &= \mathbf{z}_{TP} \times \left((-1,1,-1)^{T} \circ \mathbf{z}_{TP} \right), \\ \mathbf{x}_{TP} &= \mathbf{y}_{TP} \times \mathbf{z}_{TP}, \end{aligned} \tag{21.17}$$

where $\mathbf{n}_{TP}$ is the normal vector of reference point and plane, and $(\circ)$ denotes the Hadamard product (elementwise multiplication). Projecting an environment map point ${}^{W}\mathbf{P}_{EM}$ onto point ${}^{TP}\mathbf{P}_{TP}$ in the local tangent plane simply involves a rotation and scaling as in

$$\begin{aligned} {}^{TP}\mathbf{P}_{EM} &= {}^{TP}_{W}\mathrm{R}\, {}^{W}\mathbf{P}_{EM}, \\ {}^{TP}\mathbf{P}_{TP} &= {}^{TP}\mathbf{P}_{EM} \,/\, {}^{TP}P_{EMz}, \end{aligned} \tag{21.18}$$

where the rotation ${}^{TP}_{W}\mathrm{R}$ is defined as in

$${}^{TP}_{W}\mathrm{R} = \left[\mathbf{x}_{TP} \quad \mathbf{y}_{TP} \quad \mathbf{z}_{TP} \right]. \tag{21.19}$$

Super-resolution corneal images

Corneal reflection modeling enables a large number of applications. In reality, however, even if we manually capture images with a high-resolution camera that is placed near to the eye and carefully adjusted to avoid defocus and motion blur, the quality of corneal reflections is largely limited due to several factors. These include the low resolution of the eye region and the large field of view of the corneal mirror; the low contrast as the reflectivity of the cornea is less than 1% (Kaufman & Alm, 2003); the contamination with iris texture reflections (Wang, Lin, Liu, & Kang, 2005); and the distortions from an unknown corneal shape. In practice, therefore, high-quality corneal imaging turns out to be very challenging.

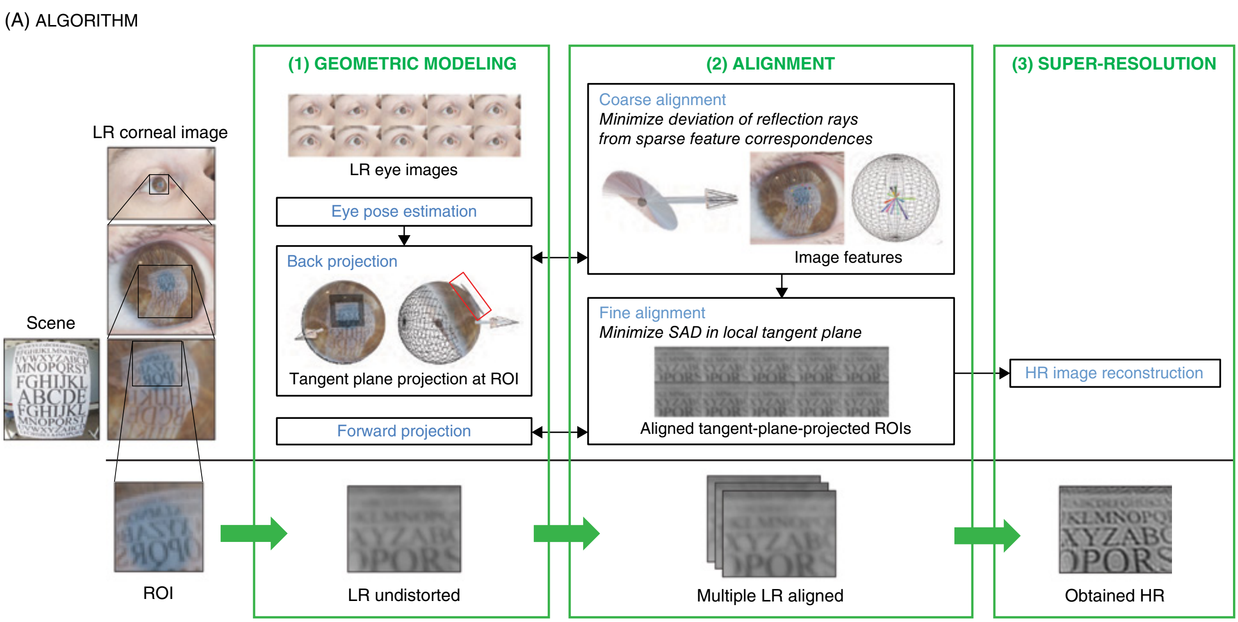
The quality of corneal images may be increased through novel developments in camera technology and special capturing systems, such as pan-tilt-zoom (PTZ) cameras (Reale et al., 2010; Yoo & Chung, 2005) and camera arrays (Chong et al., 2017) for improving resolution; light-field cameras (Lytro, Lytro Light Field imaging platform https://www.lytro.com; Raytrix, Raytrix Light Field Technology, https://raytrix.de) for improving eye-pose estimation and defocus blur; and high-dynamic-range (HDR) imaging techniques (Debevec & Malik, 1997) and cameras (El Hafi et al., 2016; ViewPLUS, ViewPLUS XViii 18bit camera, http://www.viewplus.co.jp) for improving contrast, noise, and defocus blur. This sounds promising, but the present state in these areas still involves high efforts regarding research, development, setup, and data processing. Therefore, it will take time to see practical implementations. For a more detailed discussion, see the section on image acquisition.

Focusing on conventional cameras and image acquisition, how can we increase the quality of corneal reflections? In the context of eavesdropping in information security, Backes, Dürmuth, and Unruh (2008), and Backes et al. (2009) show an approach for eye images captured at a distance. They describe a nonblind deconvolution strategy that compensates for defocus and motion blur. However, it does not solve the catadioptric distortion, nonuniform sampling, and low resolution. Motivated by the appearance of extreme high-resolution corneal imaging in science fiction (for example, *Ghost in the Shell,* produced by Motion Pictures, 2004), Nitschke and Nakazawa (2012) show an approach to overcome the issues and increase practicability of corneal image acquisition through a super-resolution (SR) strategy that reconstructs a high-resolution (HR) scene image from a series of lower resolution (LR) corneal images such as those that occur in surveillance or personal videos.

The problem is not trivial, as a standard SR algorithm (Tian & Ma, 2011) cannot be directly applied to corneal reflections. The reason is that the corneal image is the result of a nonlinear specular reflection followed by a perspective projection, which does not allow for multiple image alignment through linear transformations, and cannot obtain uniformly sampled and undistorted high-resolution scene information even from a single high-resolution corneal image. While there exist a few studies regarding central catadioptric systems (having a single focal point) (Arican & Frossard, 2011; Nagahara, Yagi, & Yachida, 2001), the cornea–camera configuration results in a dynamic noncentral catadioptric system that requires per-frame calibration through eye (corneal mirror) pose estimation.

To enable SR for corneal imaging and noncentral catadioptric imaging in general, Nitschke and Nakazawa (2012) describe an algorithm (Figure 21.10a), consisting of the following three steps:

1 Environment map reconstruction for each image.
2 Multi-image alignment by a coarse-to-fine strategy.
3 SR image estimation from the aligned images.


(A) ALGORITHM
(1) GEOMETRIC MODELING
(2) ALIGNMENT
(3) SUPER-RESOLUTION
LR corneal image
Scene
LR eye images
Eye pose estimation
Back projection
Tangent plane projection at ROI
Forward projection
Coarse alignment
Minimize deviation of reflection rays
from sparse feature correspondences
Image features
Fine alignment
Minimize SAD in local tangent plane
Aligned tangent-plane-projected ROIs
HR image reconstruction
ROI
LR undistorted
Multiple LR aligned
Obtained HR

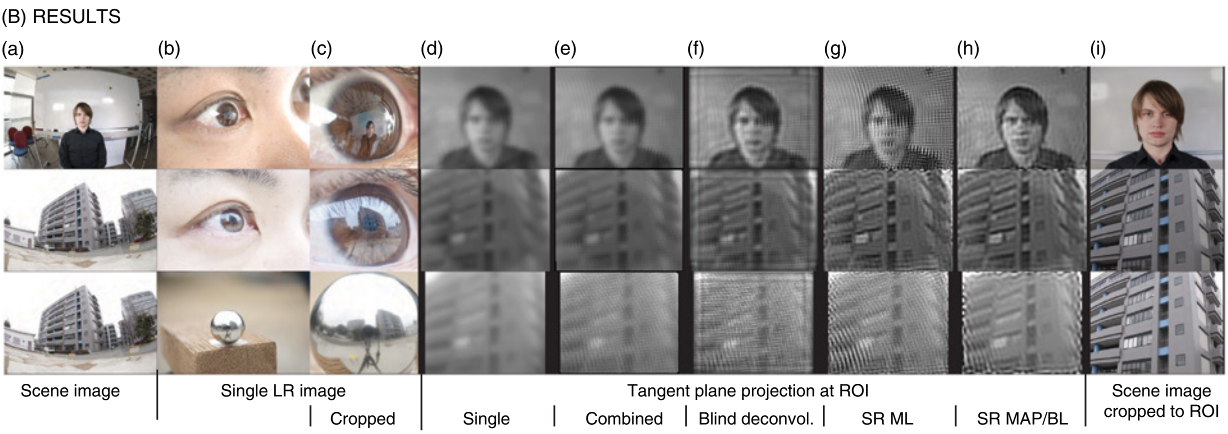


Figure 21.10 Super resolution from corneal images. (A) The algorithm involves a three-step approach comprising (1) eye pose estimation and calculation of LR environment maps, (2) registration of multiple maps and optimization of system parameters, and (3) reconstruction of an HR image. (B) SR result for different scenes (from 10 LR images). (a) Scene image. (b) Single LR image. (c) Cropped image. (d)–(g) Environment map local plane projection at region of interest (ROI): (d) Single LR image. (e) Combined aligned LR images. (f) Blind deconvolution of (e). (g),(h) SR result: (g) Maximum-likelihood (ML). (h) Maximum a posteriori (MAP) with bilateral filter residual prior (BL) (Tomasi and Manduchi, 1998). (i) Scene image, cropped to ROI. The bottom row shows results from a spherical mirror at the size of the human cornea.

Multi-image alignment. Regarding the small change of corneal locations in continuous video frames, it is feasible to assume the cornea to be centered at the world origin, where the task of alignment amounts to finding the pose of the camera with respect to the world frame. This is achieved through a multistep iterative process: (a) *Coarse alignment* is carried out using at least two feature correspondences for each LR image (Figure 21.10b–d). The transformation between the environment maps is a rotation around the origin, estimated by minimizing the deviation between the feature reflection rays. The eye pose estimation is further improved by adjusting corneal sphere locations through a pairwise and a bundle registration. (b) *Fine alignment* is carried out through image matching in the local tangent plane (Figure 21.10e), by minimizing the sum of absolute differences (SAD) at uniform sampling points using forward-projection lookup. The remaining misalignment is corrected through a 2D subpixel rigid registration in the plane.

Superresolution image estimation. For each LR eye image, the region of interest is projected to the local tangent plane, applying the recovered alignment.

The obtained (LR) points represent nonuniform samples (observations) of an unknown HR image, estimated through a MAP (maximum a posteriori) based SR approach under a Gaussian point spread function (PSF) assumption. Evaluation shows that the strategies using bilateral filter residual (MAP-BL) (Tomasi & Manduchi, 1998) or bilateral total variation filter (MAP-BTV) (Farsiu, Robinson, Elad, & Milanfar, 2004) image priors perform best and recover lost high-frequency textures (with a quality high enough to recognize small characters, human faces and fine structures) (Figure 21.10h). Similar results for a spherical mirror suggest applicability to other noncentral catadioptric systems such as specular and liquid surfaces in everyday environments.

Corneal reflection–iris texture separation

A large part of the light entering the eye diffusely reflects off the iris. This effect is largest in bright eyes, as the pigment in dark eyes absorbs most of the visible wavelength light (Proença, 2013). When capturing an image of an eye, the iris region is a mixture of the refracted iris texture and the corneal surface reflection of scene illumination. As the sources act as mutual noise, several methods are proposed to separate iris texture and corneal reflection: Wang et al. (2008) introduce a separation method that exploits the color chromaticity of iris texture from both eyes. Different approaches rely on the simple assumptions that the brightest pixels in the iris region correspond to the scene reflection, and that there exist high-contrast edges between scene reflections and iris texture. Based on these ideas, He, Tan, Sun and Qiu (2009) obtain a reflection map from an iris region using an adaptive thresholding approach and apply a bilinear interpolation to fill out the region. Tan, He, and Sun (2010) use a labeling-based corneal reflection removal for the purpose of iris segmentation. To simplify the problem and use generic image-processing techniques on a single eye image, Takemura et al. (2014a, b) argue that the effect of iris color in a corneal reflection image can be modeled as the effect of illumination with a colored light source. This allows applying computational color constancy techniques to estimate the iris color as the color of illumination, and perform chromatic adaption, so that the corneal image appears to be imaged under canonical white light. Out of the variety of color constancy algorithms (Barnard, Cardei, & Funt, 2002; Barnard, Martin, Coath, & Funt, 2002), they use the Gray-World algorithm (Buchsbaum, 1980) that assumes the deviation of the average color from grey is caused by the light source. Thus, the image is corrected by applying the transformation that aligns the average color with the intensity axis in color space. As all these approaches rely on heuristic rules, such as assuming bright scene reflections with sharp edges or consistent chromaticity in iris colors, they have low performance in scenes, where the assumptions do not hold. A promising strategy to simplify the problem and obtain a more accurate and robust solution could be to introduce explicit geometric modeling and a pixelwise registration between an eye and a scene image.

Having solved the separation problem, the recovered iris texture is important, for instance, for iris recognition in biometrics (Bowyer et al., 2008; Daugman, 2004), iris synthesis in computer graphics (Lam & Baranoski, 2006; Lefohn, Budge, Shirley, Caruso, & Reinhard, 2003), and iris texture tracking for very accurate eye-pose estimation, such as in medical applications. While the aim of this research is to remove corneal reflections from iris images, it may also be used to remove iris texture from corneal images. Modeling the eye geometry, dynamics and light interaction, and removing corneal reflections, can provide several benefits in iris recognition, such as correcting corneal refraction; supporting iris recognition "in the wild" (e.g., using conventional surveillance and personal device cameras) by illumination normalization and handling gaze angles; and identifying forged iris templates and textured contact lenses.

Corneal reflection–scene matching

Eye-image analysis has applications in many fields. However, the poor quality of corneal reflections and the overlay with iris texture causes serious difficulties for corneal-reflection and iris-texture related tasks. Finding the relation between a corneal reflection and a scene image is an important problem, as it links the eye with high-quality scene information, suitable for processing with generic computational methods. It enables problem solutions in various fields, for example. In visual recognition, this allows analysis of the scene in which a person was imaged and that triggered the person's behavior (Nishino & Nayar, 2006). In eye-gaze tracking (EGT), this allows estimating the point of gaze (PoG) in a scene from only a pair of freely captured eye and scene images (see the section on 3D point-of-gaze tracking). In virtual and augmented reality, this allows for automatic eye localization required for correctly displaying content with head-mounted or head-up displays (see the section on head-mounted display calibration). And, in iris biometrics (Bowyer et al., 2008), this may allow elimination of scene reflections from iris images to increase the reliability of the result, especially under challenging conditions with natural light and consumer devices (see the section on corneal reflection–iris texture separation). Thus, a solution to the registration problem enables eye image analysis in the field, such as from surveillance and smartphone cameras, which largely increases the number of potential application scenarios. In general, the registration of body features (such as face and eye features), and direct and reflected scene features, across multiple eye and scene images, in combination with a suitable model, can help to increase accuracy, solve for unknown parameters, and even target novel application scenarios.

While image registration is a common problem with existing algorithms, the application to corneal images is challenging due to the large amount of noise from iris texture, eyelid and eyelash occlusion, and the low contrast and the nonlinear geometric distortion from the reflection at the curved corneal surface. The major differences to conventional image registration are the nonlinearity of the problem and the difficulty to obtain dense correct correspondence pairs due to the large nonsimilarity of the images. In the following, we review two approaches that have been developed to solve the task of robust correspondence matching: (a) active-light methods that augment the scene with coded illumination, which can be recovered from within an eye image and a scene image (Nakazawa & Nitschke, 2012; Nitschke, 2011), and (b) passive methods that rely on geometric modeling and image correspondence matching (Nakazawa et al., 2016; Takemura et al., 2014a). Regarding eye and scene image analysis, robust direct correspondence matching enables for several benefits such as the absence of geometric eye–scene calibration, and the support of arbitrary depth-varying scenes and dynamic interactive scenarios (with remote and head-mounted systems). Moreover, it enables the computation of eye-related information from conventional images with much higher quality.

Active-light methods Coded structured light techniques allow for robust optical information transmission (Kagami, 2010) and correspondence matching between images, scene locations and light emitters (commonly digital video projectors) (Salvi, Fernandez, Pribanic, & Lado, 2010). In the context of corneal imaging and eye-gaze tracking, Nakazawa and Nitschke (2012), Nitschke (2011), and Nitschke, Nakazawa, and Takemura (2011b) use coded illumination for robust correspondence matching between eye reflection and scene images, which is difficult to achieve with passively captured images through natural feature tracking and epipolar geometry. The strategy benefits from accuracy and precision through pixelwise image-based matching, and robustness to challenging scene conditions and eye appearance.[6] For further details, see the section on 3D point-of-gaze tracking.

In practice, nonintrusiveness through invisible, imperceptible, and removable structured light (Fofi, Sliwa, & Voisin, 2004) is important to avoid affecting the scene appearance for human observers and to allow removal from camera images for the recovery of visual scene information. Invisible structured light operates in wavelengths outside the visible spectrum, where infrared (IR) light is mainly applied as it can be detected with common imaging sensors. Since the lamps of digital video projectors are optimized to minimize the emission of IR light, the standard lamp may be replaced with a special IR light source (Lee, Hudson, & Dietz, 2007). Another technique is imperceptible structured light, where a sequence of alternating light and complement patterns is projected using visible light. If the sequence exceeds the critical flicker frequency (CFF) (Watson, 1986, pp. 6-1–6-43), the dynamic content is visually integrated over time and perceived as a static illumination. Although it is imperceptible for human observers, the pattern is detected in images from a synchronized camera.

The concept is verified by three prototype implementations, using either (a) a standard digital video projector (Nitschke, 2011), (b) a specially designed LED-array projector (LED-AP) to illuminate the scene (Nakazawa & Nitschke, 2012), or (c) specially designed LED markers that are directly attached to scene locations (Nakazawa, Kato, Nitschke, & Nishida, 2017) (Figure 21.11). An eye camera takes a close-up view of the eye region, where the reflected patterns are observed and decoded. IR LEDs attached to the eye camera allow for pupil segmentation and

[6] Special LED arrays have been proposed for automatic reflection extraction and increased robustness in eye-gaze tracking—Hua, Krishnaswamy, & Rolland, 2006, and Li, Kolakowski, & Pelz, 2007. Although termed "structured light," this relates only to the geometric alignment, not the encoding of light.

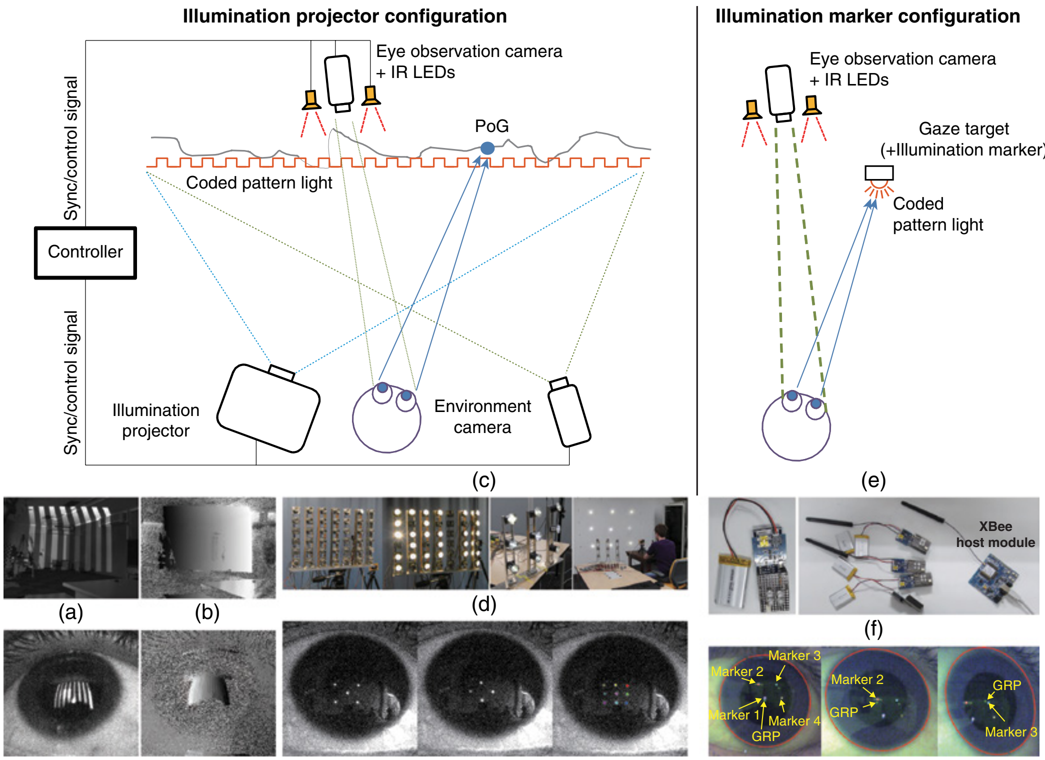


Figure 21.11 Active-illumination devices for eye–scene matching in the context of point of gaze (PoG) tracking. (left) Illumination projector configuration. A user is gazing an unknown PoG on an arbitrary surface. An eye camera tracks a close-up image of the user's eye. An environment camera captures a view of the gazed surface. Coded structured light is projected onto the surface, from where it reflects towards environment camera and eye, from where it again reflects into the eye camera. Correspondences are obtained by decoding the reflections in both images. (a)–(b) Off-the-shelf digital video projector setup with alternating-pattern imperceptible projection: (a) Single binary-coded *x*-coordinate code-bit image for scene and eye camera, and (b) decoded *x*-coordinate from a time series of code-bit images. (c)–(d) LED array projector (LED-AP) featuring direction-adjustable, high-frequency and high-power IR-LED units. (c) 42-LED and 9-LED projectors with visible light LEDs. (d) Two binary-coded code-bit images and decoded correspondence IDs (indicated by color/intensity). (right) Illumination marker configuration. (e) Markers, featuring the same high-frequency and high-power IR-LED units, are attached to particular scene locations (gaze target objects). Marker illumination and eye camera image acquisition is synchronized using an XBee wireless network. (f) Calculating the gaze reflection point (GRP) and decoding the marker IDs (indicated by color/intensity) in an eye image allows identifying when the user looks at a particular gaze target.

eye-pose estimation. In the projector configuration, correspondences are obtained by matching codes in the eye and scene image. In the marker configuration, correspondences are obtained directly by identifying particular markers in the eye image. In all configurations, illumination and image acquisition is synchronized.

Digital video projector. Using a standard data projector is the easiest way to set up the system. Beside using off-the-shelf hardware, the approach combines several advantages, including high resolution for dense correspondence matching, high intensity for increased robustness to environmental light, and short-focus lenses for wide-scene coverage. Nevertheless, there are also disadvantages for practical application, including relatively low frame rates for temporal code projection and low IR light intensity for invisible projection. A proof-of-concept implementation uses a 10-bit binary Gray code to encode *x*- and *y*-coordinates with imperceptible pattern projection, which results in a sequence of 40 frames for 1 effective frame (Figure 21.11(left), a, b).

LED array projector. To compensate for the disadvantages of standard data projectors, Nakazawa and Nitschke (2012) introduce a programmable LED-array projector (LED-AP) that produces high-power and high-frequency light beams to achieve real-time correspondence matching. It consists of an array of individually modulated adjustable IR-LED lens units, LED power modules and a controller (Figure 21.11(left), c, d). A linear interpolation technique compensates for the relatively low resolution (number of LEDs) as compared to digital video projectors, where dense correspondences are obtained with respect to four neighboring light spots. The temporal encoding uses either a binary numeral code or the Hamming(7,4) code (Moon, 2005). The binary numeral code is simple and short length, however, does not have error compensation functionality. The Hamming(7,4) code encodes 4-bit base information into a 7-bit length code, and allows to correct a 1-bit error and to detect a 2-bit error in one code block. In practice, this allows to correct / discard wrong detections due to occlusion or reflection.

LED marker. Smith, Vertegaal, and Sohn (2005), and later Nakazawa and Nitschke (2017) introduce a coded illumination marker, designed to be attached onto target objects in the scene. The latter consists of an Arduino microcontroller, an XBee wireless module for camera synchronization, and IR-LEDs that emit unique temporal-coded patterns (Figure 21.11(right), e, f). The advantage of this configuration lies in its simplicity, since the scene correspondences are recovered directly from an eye camera image without the requirement for a scene camera. The temporal encoding uses a modified binary code for the purpose of robust detection, where two additional frames with all markers turned off and on allow to detect the marker positions.

Passive methods While active illumination provides dense and robust registration through simple image processing, it also limits to complex setups and controlled, often static and indoor environments. Several approaches exist towards the passive registration of eye and scene images.

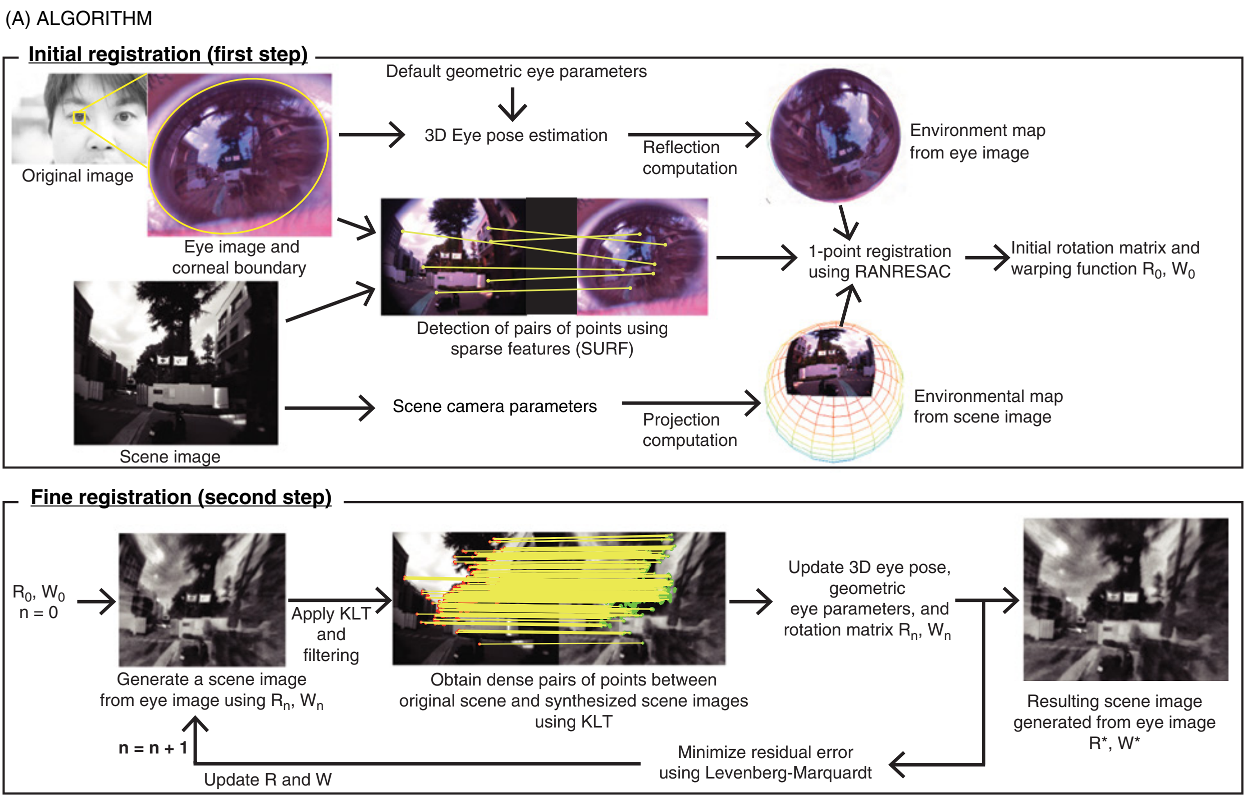
Epipolar geometry. Nishino and Nayar (2006) discuss the correspondence matching between multiple eye images for the purpose of 3D scene reconstruction (see the sections on scene point position estimation and display–camera calibration). They define the epipolar geometry of the corneal stereo system formed by the two eyes of the same face image, which naturally extends to more than two eyes in multiple face images through eye–camera tracking. While the epipolar geometry reduces the correspondence-finding problem from the whole corneal image to the epipolar curve, automatic matching suffers from the large iris texture noise. For a proof-of-concept evaluation of 3D reconstruction, they circumvent the problem by manually specifying object correspondences along the epipolar curves in the two eyes of a face image.

Straightforward correspondence matching. Takemura et al. (2014b) describe a manual strategy for evaluation of simple object recognition in corneal images. The study targets a gaze-based guidance system, where the objects are eight planar outdoor direction boards of same shape and size. The corneal images are extracted through projection into a tangent plane at the optical axis (see the section on local tangent plane projection) and warping to rectify the imaged boards (using a homography matrix estimated from the four manually selected board corners). Then the correct object match is detected as the one resulting in the lowest accumulated SIFT (Lowe, 2004) correspondence error between the warped corneal image and an object image. Takemura et al. (2014a) improve on this to obtain an automatic online strategy for gaze mapping and object recognition in a scene camera image. Automatic mapping is achieved by replacing the manual homography definition with template matching between corneal image and scene camera image. Increased gaze-mapping accuracy is achieved through tangent-plane projection at the individually calibrated visual axis instead of the optical axis. The strategy is a large improvement over previous works, but still suffers from two major issues: (a) Despite basic iris color compensation through color constancy, the matching may still suffer from iris texture noise; and (b) compared to sparse feature matching, dense template matching needs a good initial guess and is computationally expensive.

Robust correspondence matching. To overcome these issues, Nakazawa et al. (2016) propose a generic coarse-to-fine registration method for noisy imagery that is shown to successfully align an eye and a scene image. The approach achieves robustness through a random-sampling based verification scheme that requires only a single correct correspondence pair, and accuracy through iterative dense correspondence matching and geometric modeling.

The two-step strategy works as follows (Figure 21.12A): First, the initial registration applies a problem simplification, assuming the common configuration, where the distance to the scene is much larger than the distance between the cameras. Then, the complex nonlinear 2D warping problem reduces to determining the 3D rotation that aligns the spherical environment maps from corneal reflection and scene image. The two unknown rotation parameters can be recovered from a single correspondence pair of standard rotation-invariant local image features (Tuytelaars & Mikolajczyk, 2008), such as SIFT (Lowe, 2004), SURF (Bay, Tuytelaars, & Gool, 2006) and MSER (Matas, Chum, Urban, & Pajdla, 2002), which encode both, the texture position and orientation. To robustly determine at least a single correct (inlier) correspondence pair, a *RANRESAC* (RANdom RESAmpling Consensus) strategy (Nakazawa, 2016) is developed. RANRESAC verifies each registration hypothesis (obtained from a correspondence pair) through generating new random correspondences according to the warping function. In contrast to RANSAC (RANdom SAmpling Consensus) (Fischler & Bolles, 1981) and its extensions, this scheme does not assume the majority of initial correspondence pairs to be inliers. It can therefore be robustly applied to noisy images, where it is difficult to obtain multiple correct pairs.

Due to modeling and estimation errors, the sparse registration achieves only reasonable accuracy, decreasing with distance from the correspondence location(s). To account for this, the second step performs an iterative fine registration that



(A) ALGORITHM
Initial registration (first step)
Original image
Eye image and corneal boundary
Default geometric eye parameters
3D Eye pose estimation
Reflection computation
Environment map from eye image
Detection of pairs of points using sparse features (SURF)
1-point registration using RANRESAC
Initial rotation matrix and warping function R_0, W_0
Scene image
Scene camera parameters
Projection computation
Environmental map from scene image
Fine registration (second step)
R_0, W_0
n = 0
Generate a scene image from eye image using R_n, W_n
Apply KLT and filtering
Obtain dense pairs of points between original scene and synthesized scene images using KLT
Update 3D eye pose, geometric eye parameters, and rotation matrix R_n, W_n
Resulting scene image generated from eye image R*, W*
Minimize residual error using Levenberg-Marquardt
n = n + 1
Update R and W

Figure 21.12 Robust registration of corneal reflection and scene images. (A) The algorithm consists of an initial and a fine registration step. To turn the task into a tractable problem, the complex nonlinear image warping is modeled as the rotation of the spherical environment maps (EM) of corneal reflection and scene image. The rotation can be parameterized from a single local feature correspondence pair. In the initial step, at least one correct (inlier) pair is determined through the one-point algorithm that applies a RANRESAC strategy to verify registration hypotheses. Unfortunately, the result is not accurate for the entire image region due to (1) the sparsity of the feature points, (2) the fixed geometric eye parameters, and (3) the error in eye pose estimation. To solve these problems, the fine registration step performs an iterative joint optimization of EM rotation, eye pose and eye model parameters using additional dense correspondence matches. (B) Evaluation results. (a) Cropped eye images indicating the gaze reflection point (GRP) (reflection of the point of gaze (PoG)). Registration allows (b) mapping the GRP from an eye image to the PoG in a scene image, and (c) warping the whole scene reflection into the view of the scene image. This enables various applications, such as (d) mapping of a gaze trajectory from a head-mounted eye camera into a stationary scene camera or Google Street View imagery, and (e) mapping of the whole visual field for analyzing peripheral vision.

optimizes the initial estimate by searching for environment map rotation, eye pose and eye model parameters. The procedure minimizes the residual error of dense correspondence matches from the optical-flow between the scene image and a synthesized image from the corneal reflection. This second step introduces two major benefits: (a) The dense optical-flow tracking achieves an accurate registration for the entire image region, which cannot be represented though the catadioptric image formation alone. (b) The procedure also optimizes eye model and pose parameters, which is beneficial for the target applications.

An additional improvement is achieved by introducing an aspherical shape and reflection model for the cornea. However, finding optimal asphericity parameters is challenging, as a straightforward joint estimation is not possible due to (a) a dependency on other parameters such as eye pose and distance, and (b) a sensitivity to the tracking errors in the images. Thus, the effectiveness of the model is verified in a separate step for a set of discrete asphericity values. As a result, evaluation with four subjects under five indoor and outdoor conditions achieves a mean error of 3.37°, 1.27° and 1.05° for initial registration, fine registration and fine registration considering asphericity, respectively.

Nakazawa et al. (2016) further show two applications in eye-gaze tracking that cannot be achieved with existing systems, namely (a) accurate eye pose estimation without the need for IR illumination, and (b) ad hoc point-of-gaze (PoG) and peripheral vision tracking under an uncalibrated dynamic eye to scene relation. First, PoG estimation from uncalibrated eye and scene images is realized by combining corneal-imaging-based PoG estimation (Nitschke & Nakazawa, 2012) with the described image registration. Such a technique allows for more flexible configurations of devices, scenes, and use cases. They demonstrate a setup consisting of a head-mounted eye camera and a ground-mounted fisheye scene camera (Figure 21.12d), where the latter may also be replaced with an existing scene image database such as Google Street View. Second, peripheral vision mapping is realized by combining corneal reflection modeling with the pixelwise warping function obtained from the registration (Figure 21.12e). The ability to track the complete visual field provides the potential for a new generation of EGT systems. For more details, see the section on 3D point-of-gaze tracking.

The registration method not only allows for system simplification, but also extends application scenarios to ad hoc and long-term use as well as postprocessing of data from unknown setups. It could have a significant impact on several important problems, including visual saliency estimation, driver's view analysis and psychological tasks, and can enable more accurate and effective solutions as well as a deeper understanding of human vision. Furthermore, it is not limited to eye images. The described modeling and verification scheme allows application to the robust registration of noisy images in general, where RANSAC-based approaches usually fail. Potential application scenarios include dirty lenses, traffic mirrors, and reflecting windows.

Applications

The previous sections introduced a methodical framework for corneal reflection modeling from multiple images that recovers the geometric relationship between cameras, eyes and scene. Removing the requirement for an explicit geometric calibration allows existing limitations to be overcome and enables novel approaches in visual eye analysis.

For example, it naturally supports dynamic scenarios, such as in user tracking, mobile systems, and changing environments; and scenarios where a dedicated calibration cannot be applied—because of time, ability, or awareness constraints. In summary, this can remove the need for complex hardware and setup procedures, scene restrictions, expert supervision and subject awareness.

We believe that this has the potential to facilitate a wide range of applications and, in this section, want to discuss general implications. After that, we will review three major applications from different fields, comprising (a) 3D scene reconstruction to estimate the pose of a computer screen in computer vision, (b) eye-pose estimation to calibrate head-mounted and head-up displays in augmented and virtual reality, and (c) nonintrusive calibration-free eye gaze tracking (EGT) for everyday environments in HCI.

General implications

Visual recognition. The environment map from a corneal image provides information about the location and situation in which a person was photographed (Nishino & Nayar, 2006). The result allows for scene information mining in surveillance (Nitschke & Nakazawa, 2012; Motion Pictures, 2004), eavesdropping (Backes et al., 2009), and forensics (Jenkins & Kerr, 2013; Johnson and Farid, 2007; Motion Pictures, 2004), and provides context information in face / body analysis and scene understanding. A wearable corneal imaging camera (Lander, Krüger, Löchtefeld, 2016; Nakazawa et al., 2015; Takemura et al., 2014b) with automated real-time processing can be leveraged in life logging for personal assistance and diary keeping, and contribute to ambient intelligence systems such as in driver assistance and smart home. Apart from computational analysis, low-quality corneal reflections successfully allow for human analysis in identity discrimination, familiar face recognition, and even emotional state and interest-recognition tasks (Jenkins & Kerr, 2013).

Computer graphics and vision. Besides the visual information of the scene, the environment map also provides directional information about the incident illumination at the eye, which allows for recovery of scene reflectance and structure. This facilitates a number of applications in computer graphics and vision, such as scene relighting (Nishino & Nayar, 2004; Tsumura et al., 2003), illumination normalization in object, face (Nishino et al., 2005) and iris recognition (Daugman, 2004), and illumination consistency analysis in digital forgery detection (Johnson & Farid, 2007). If multiple eye images from different viewpoints and scene images are available, it may be possible to reconstruct a 3D scene model or improve the quality of corneal reflections using corneal image-processing techniques. This further allows for applications, for example in HCI (Hansen & Ji, 2010; Lambooij, IJsselsteijn, Fortuin, & Heynderickx, 2009), and augmented and virtual reality (AR/VR) (Itoh & Klinker, 2014; Plopski, 2016; Plopski et al., 2015a). In the fields of human science, medicine, biology, and biometrics, the appearance of eye features may be of concern, and superimposed corneal reflections would act as noise. Here, a combination of scene structure and illumination could provide constraints for removing scene reflections and correcting for corneal refraction.

Human–computer interaction. Eye tracking is a common task with a long history (Duchowski, 2007; Hansen & Ji, 2010; Holmqvist et al., 2011; Young & Sheena, 1975) that is relevant to a large number of applications in a variety of fields (Duchowski, 2002). Nevertheless, there still remain major issues to be solved. Eye tracking is traditionally applied to the gaze-based interaction with controlled planar screens (Bolt, 1982). Exploiting the discussed geometric eye modeling and tracking techniques, however, provides additional information and flexibility that can enable application for intelligent sensors in ubiquitous and ambient scenarios, integrated with practical implementations in off-the-shelf products. Introducing anthropometric knowledge permits robust passive eye pose estimation, and tracking of the complete perceived field of view of a person (instead of only the point of gaze (PoG)) (Nakazawa et al., 2015; Nishino & Nayar, 2006). In conclusion, the gained support for uncalibrated dynamic scenarios and arbitrary environments (Nakazawa et al., 2016; Nakazawa & Nitschke, 2012; Nitschke, 2011; Nitschke et al., 2013a; Takemura et al., 2014a, b) allows for novel applications areas, such as human factors engineering, human behavior and intent analysis, and human–robot interaction (HRI). Especially applications with infants and children, elderly and even nonhuman subjects (Machado & Nelson, 2011; Somppi, Törnqvist, Hänninen, Krause, & Vainio, 2012) can benefit from nonintrusive remote observation through the absence of required expertise and obtrusive body attachments (Chong et al., 2017).

Diagnostic studies. Capturing a person's eyes and face in the same image allows joint analysis of the gazed scene together with the person's reaction (Matsumoto, Keltner, Shiota, O'Sullivan, & Frank, 2008). Such stimulus–response information enables diagnostic studies in different disciplines, including medicine, psychology, engineering and marketing. Specifically, this can help to diagnose degrading of the visual and motor systems, analyze human factors, or understand the human mind, attraction, problem solving, communication, interaction and social networks. Recently, the combination of corneal imaging and gaze behavior is studied for the early diagnosis of autism spectrum disorders (ASD) in infants (Chong et al., 2017). Geometric modeling provides the following additional benefits compared to traditional eye analysis: (a) Remote visual inspection facilitates nonintrusive methods, where equipment and conditions do not interfere with the task or otherwise affect the subject. (b) The absence of geometric calibration allows for interactive studies and unknown conditions that can provide novel insights. (c) Correspondence matching between corneal reflections and direct views provides high-quality scene information for further processing.

Display–camera calibration

With advances in vision algorithms, the webcam moves beyond solely being a tool for videoconferencing. Together with a standard monitor or projection screen, it forms a display-camera system. In the past, there have been two major areas of application: One is the reconstruction of 3D object properties, such as shape and reflectance, where the display is used as a controlled planar light source to illuminate the scene captured by the camera. The other is HCI, where the content of the display is adapted according to information about the user obtained from the camera. In object reconstruction, photometric stereo methods are used to estimate the shape of Lambertian (Clark, 2010; Funk & Yang, 2007; Schindler, 2008) and partially Lambertian objects (Francken, Hermans, Cuypers, & Bekaert, 2008). However, the nonfocused and nondirectional display illumination is ideal for reconstructing non-Lambertian specular (Bonfort, Sturm, & Gargallo, 2006; Francken, Cuypers, Mertens, Gielis, & Bekaert, 2008; Tarini, Lensch, Goesele, & Seidel, 2005), and transparent / translucent objects (Kutulakos & Steger, 2008; Morris & Kutulakos, 2007). Vision-based user interfaces employ computer vision to "look at people" and perform tasks such as body (parts) detection and tracking, and face, facial expression, gesture, posture, activity and behavior recognition (Jaimes & Sebe, 2007; Porta, 2002; Turk, 2004). Display-based interactive eye-gaze tracking (EGT) (Hansen & Ji, 2010) is especially important.

The majority of applications requires geometric or photometric calibration of display and camera. Here, we focus on geometric calibration to recover the pose of the display with respect to the camera (Bonfort et al., 2006; Francken, Hermans, & Bekaert, 2009; Funk & Yang, 2007; Tarini et al., 2005). If the screen is visible in the camera image, the calibration can be performed using standard techniques detecting screen patterns (Hartley & Zisserman, 2003; Kaehler & Bradski, 2015). This is, however, not possible in the common case where screen and camera face a similar direction. Such a configuration requires moving specular objects with known shape and pose (Sturm & Bonfort, 2006; Kumar et al., 2008), such as planar (Bonfort et al., 2006; Funk & Yang, 2007) and spherical mirrors (Francken et al., 2009; Tarini et al., 2005), to reflect the screen content into the camera. The process is cumbersome as it involves a special mirror and tedious physical efforts.

Motivated by the discovery that screen reflections are clearly visible in eye images, Nitschke (2011); Nitschke et al. (2009a, b, 2011) developed a method that reconstructs the display from at least two eye images (Figure 21.13). The basic algorithm estimates 3D marker locations on the display by triangulating light paths from marker reflections under varying eye poses (see the section on scene point position estimation). Experimental evaluation, however, shows a large absolute error and deviation due to the unknown geometry and size of the individual eye. To compensate for this, a nonlinear optimization framework performs bundle adjustment of eye poses, reflection rays and reconstructed scene, subject to the geometric constraints: display planarity, size, and ray distance. Moreover, this allows for automatically resolving the inherent ambiguity that arises in passive iris-contour based eye pose estimation. Thorough experimental evaluation, using synthetic and real data, shows considerable and stable improvement with respect to varying subjects, scene poses, eye positions, and gaze directions. It further shows that the strategy also improves results obtained with a spherical mirror. The findings provide general insight into catadioptric and corneal scene reconstruction from multiple images and show that constraints can greatly improve the results, to a level that should be sufficient for many applications. An extension of this work (Nitschke et al., 2011b) aims to enhance the approach under conditions of practice by encoding display correspondences into coded illumination patterns for automatic and robust detection.

In conclusion, the method provides several advantages over previous approaches, as it (a) does not require additional hardware, (b) user interaction or awareness, (c) supports dynamic setups, and (d) estimates eye poses for eye-gaze tracking applications.

Head-mounted display calibration

Augmented reality (AR) aims at the integration of the physical world with computer-generated multimodal content, such as sound, video, graphics, and olfactics, to enhance human perception of reality. The quality of a visual AR experience depends

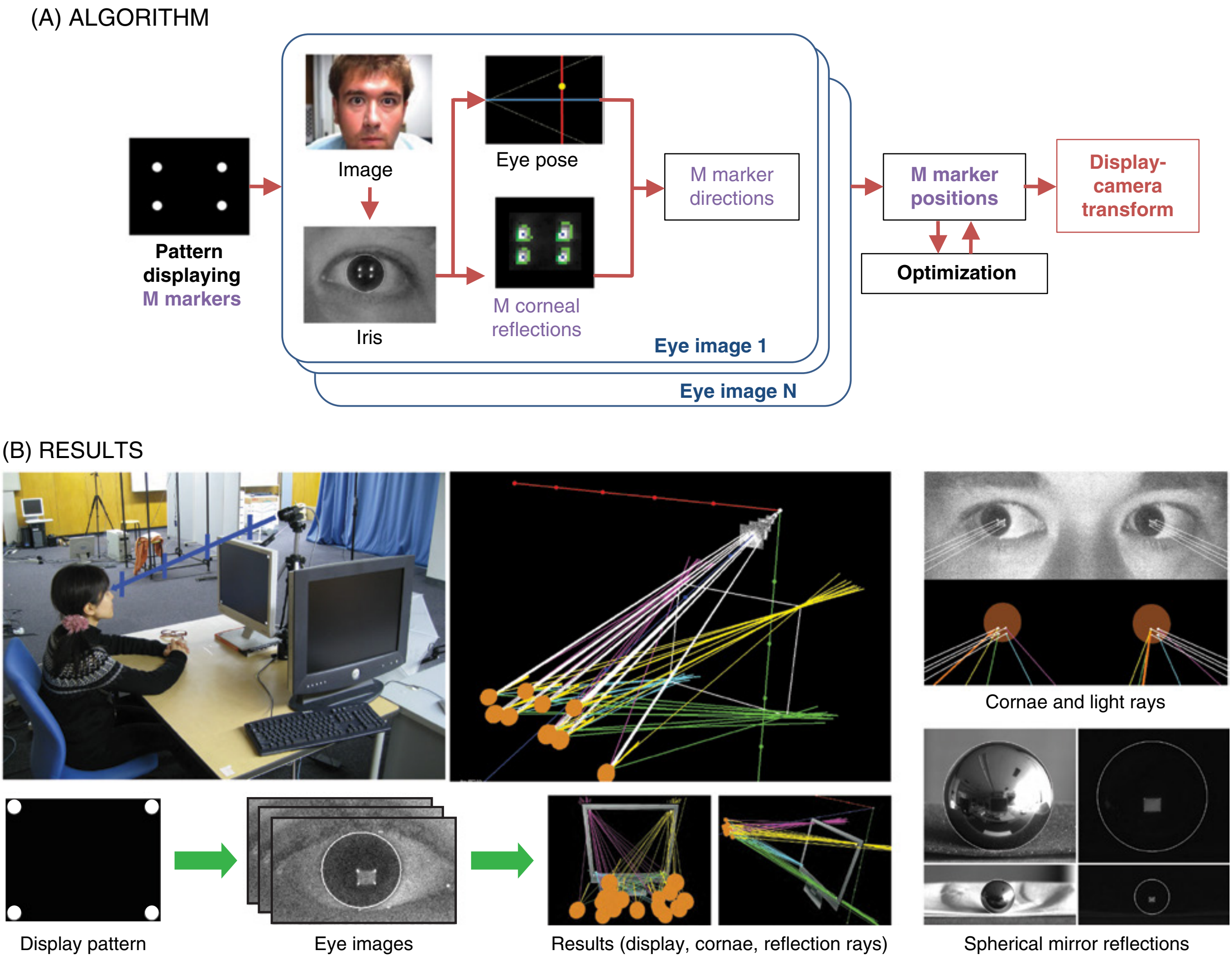


Figure 21.13 Display–camera calibration from eye reflections. (A) Algorithm: The screen shows a pattern with M circular markers. The corresponding 3D points are reconstructed from the intersections of the M corneal reflection rays under N eye poses. (B) Results, showing a typical setup with recovered display, corneal spheres (brown), back projection (white), and reflection rays (colored).

on how well virtual content is integrated into the real world—spatially, photometrically, and temporally. In other words, if the world does not appear consistent, there is a high chance that users dislike or reject the AR experience as a whole. A large number of use cases, such as maintenance, training, and medical tasks, require consistent spatial visualization, where correct alignment is of utmost importance for the ability to correctly view an augmentation (Holloway, 1997) (Figure 21.14a).

A common solution to this problem is to approximate the user's view through a scene camera that is rigidly attached to the display device, and dynamically tracked. The user is then shown an augmented view from the view point of the camera. However, this can greatly differ from the user's perspective, such as with a mobile phone or tablet PC. Recent research has shown that, instead, displaying the augmentation from the user's perspective has the potential to improve perception and overall experience (Liestol & Morrison, 2013). Head-mounted displays (HMDs) are therefore likely to be the better choice for spatially consistent visualization.

For a long time, HMDs were limited to laboratory and industrial use cases. Recent advances in display, computing and application technology, however, lead to an increasing number of inexpensive, high-quality, consumer-oriented devices. Examples are the video see-through adaptations of Oculus Rift (Oculus, Oculus Rift, https://www.oculus.com) (VST-HMDs) or optical see-through Google Glass (Google, Google Glass, https://x.company/glass/), Epson Moverio (Epson, Epson Moverio, http://www.epson.com/moverio) and Microsoft HoloLens (Microsoft, Microsoft HoloLens, http://hololens.com/) (OST-HMDs). Although OST-HMDs have a large application potential and were at the forefront of AR research (Caudell & Mizell, 1992; Rolland, Holloway, & Fuchs, 1995) they have been replaced by handheld devices and VST-HMDs due to their limitations, including small augmentable field of view, low contrast, and requirement for constant recalibration (Kishishita et al., 2014).

Design issues, like field of view and contrast, can be solved through improvements and technical progress. Spatial alignment, on the other hand, remains a challenging issue as it needs to be solved for each user and executed through a calibration process. Additionally, if the HMD moves on the user's head, the alignment is no longer ideal and the calibration needs to be repeated. Common solutions such as the Single Point Active Alignment Method (SPAAM) require extensive user input, which is tedious and introduces user-dependent errors (Axholt et al., 2011; Tuceryan, Genc, & Navab, 2002). As a result, the recalibration is often skipped, which in return impacts the user experience and acceptance of OST-HMDs.

Recently, eye tracking has become available with commercial VST-HMDs (Fove, Inc., https://www.getfove.com; SensoMotoric Instruments (SMI), Eye Tracking Platform for Virtual Reality HMDs, https://www.smivision.com) and OST-HMDs (SensoMotoric Instruments (SMI), Eye Tracking Platform for Augmented Reality HMDs and Smart Glasses, https://www.smivision.com). Several studies propose the integration of eye tracking with HMDs for user interaction and behavior analysis (Bulling, Roggen, & Tröster, 2009; Ishiguro, Mujibiya, Miyaki, & Rekimoto, 2010; Park, Lee, & Choi, 2008). However, the eye also plays an essential role in the calibration of an OST-HMD, as it defines the viewpoint for rendering virtual content. Thus, an eye tracking-equipped device allows for automatic calibration (Figure 21.14b, e). Itoh and Klinker (2014) build on this idea and propose the INteraction Free DIsplay CAlibration (INDICA) method that uses a standard webcam for passive iris contour-based eye position and gaze estimation. Additionally assuming uniform eye parameters

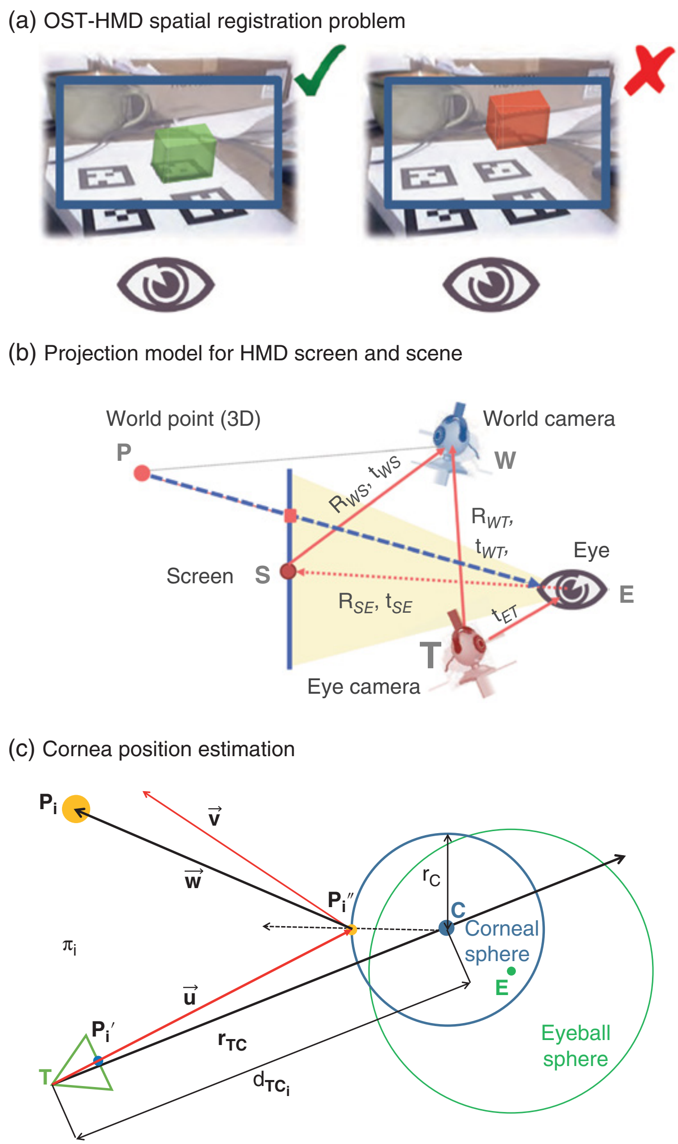


Figure 21.14 Corneal-imaging calibration (CIC) for optical see-through head-mounted displays (OST-HMDs) (Plopski et al., 2015a, and Plopski, 2016). (a) OST-HMD spatial registration problem showing the difference between aligned and nonaligned virtual content. (b) The projection is a function of HMD parameters and eye pose. Assuming a known HMD, the eye pose is obtained from the corneal reflection of the screen without requiring IR-illumination. (c) The cornea position is estimated from a single eye image with detected reflections from at least two screen locations. (d) The eyeball position is estimated from at least three images of a moving eye, as the center of the spherical orbit formed by the cornea centers. This implicitly achieves eye tracking. (e) OST-HMD with world and eye tracking camera attached. (f) Cornea position estimation from reflected screen checkerboard corners. Reprojection errors from iris contour (INDICA) (Itoh & Klinker, 2014)) and corneal-imaging tracking (CIC). (g) Eye position stability: CIC converges to an error of <1 mm within seven frames. (h) Projection errors for manual calibration (SPAAM), with degradation (Degraded SPAAM) and automatic calibration (CIC). While SPAAM achieves the lowest error, it suffers from incorporating inaccuracies, which reduces the perceived quality (Moser et al., 2015).

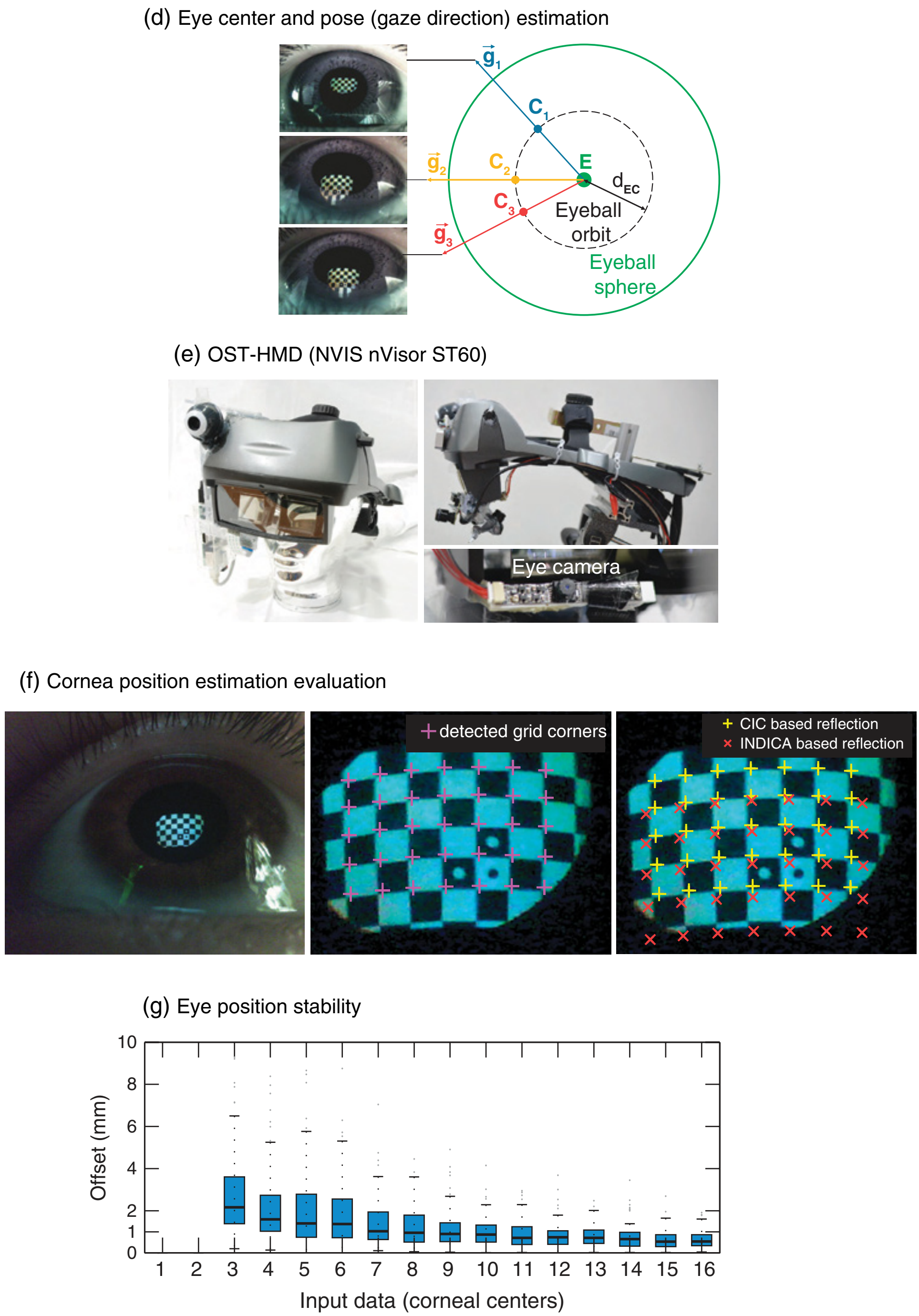


Figure 21.14 (Continued)

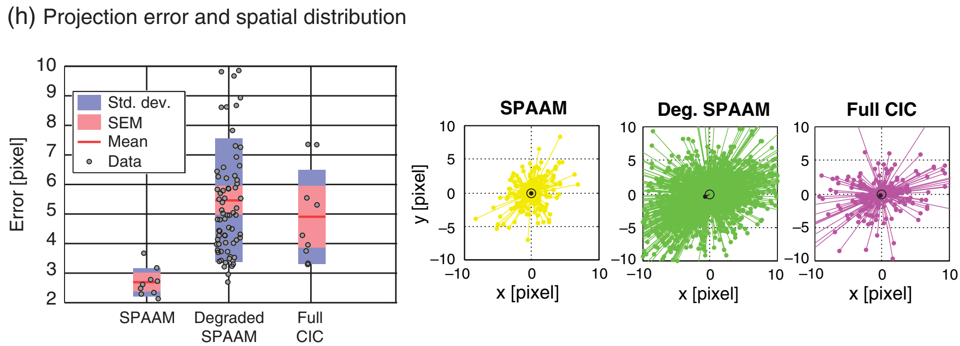


Figure 21.14 (Continued)

removes the requirement for manual calibration. However, iris contour-based eye tracking is a difficult problem, due to several factors such as occlusions from eye lids and eye lashes, a varying contrast and a gradual transition between iris and sclera, and an intrinsic ambiguity in the sign of the 3D iris tilt angle. Small tracking errors also propagate to large errors in the estimated eye pose. Methods, therefore, generally result in a large error of about 6° (Nishino & Nayar, 2006; Nitschke et al., 2013b; Wu et al., 2005), which consequently leads to low-quality eye-pose estimation (Schnieders et al., 2010). A more robust and accurate eye tracking approach is the pupil-center–cornea-reflection (PCCR) method (Guestrin & Eizenman, 2006; Hansen & Ji, 2010; Shih et al., 2000; Villanueva & Cabeza, 2008) used in commercial systems. It employs active IR illumination and synchronized image acquisition, to estimate the position of the cornea from specular highlights (glints) and the gaze direction from the segmented pupil contour. While being robust and accurate, the disadvantages are the system complexity regarding hardware and processing, problems in environments with strong illumination, and intrusiveness, with an unknown long-term impact from the exposure to IR illumination.

To compensate for this, Plopski et al. (2015a) and Plopski (2016) propose corneal imaging calibration (CIC), which combines the geometric advantages of PCCR with the properties of HMDs. Instead of using additional light sources, complex illumination and pupil tracking, the method directly tracks the corneal reflection of the screen content. It works as follows. In a preprocessing step, the HMD is calibrated to obtain the relative poses of screen, world camera, and eye camera. At runtime, first, the position of the cornea is estimated from correspondence matches of screen and corneal reflection locations in an eye image. Then, the position of the eye is obtained as the center of the rotation orbit defined from at least three noncoplanar cornea positions (Figure 21.14c, d, f). Plopski et al. (2015b) and Plopski (2016) improve the eye tracking by introducing a single-frame hybrid eye-tracking technique, combining active-light corneal image template matching with passive iris-contour tracking. It provides increased accuracy through hybrid geometric modeling with individual eye parameter estimation and increased practicality through continuous tracking of generic application content instead of artificial checkerboard patterns. The technique tracks the eye pose relative to the display, assuming a fixed precalibrated eye–camera pose. In practice, this turns out to be a limiting factor, as the eye camera often needs to be readjusted for different users and even different poses of the HMD. To account for this, Plopski et al. (2016) describe a method to determine the eye camera and HMD poses automatically by tracking corneal reflections of additional light sources, rigidly attached to both devices.

In conclusion, the corneal-imaging-based eye tracking achieves a practical and lightweight HMD (re)calibration and gaze-tracking method, where the point of gaze (PoG) is readily obtained as the intersection of the gaze ray with the screen and the scene. The approach combines several advantages: It is practical as it uses simple and automatable image processing. It is less reliant on eye modeling as the error propagates less into the result. And, it is more robust as it allows for a large number of dense subpixel correspondence matches. Thus, while being lightweight and simple, the method is suitable for practical and high-quality eye-gaze tracking (EGT) (Figure 21.14g, h).

3D point-of-gaze tracking

Eye-gaze tracking (EGT) is the problem of tracking the gaze direction (pose of the eye) or the point of gaze (PoG) in the scene (Holmqvist et al., 2011; Duchowski, 2007; Hansen & Ji, 2010; Young & Sheena, 1975). Eye-pose estimation reveals the gaze direction as the optical axis, and is, therefore, equivalent to the first part in geometric eye-gaze tracking. Point-of-gaze detection further requires calculating the intersection of the gaze ray (optical axis or visual axis, considering the individual offset) with a 3D scene model. However, especially in dynamic scenes and mobile scenarios, a 3D scene model and pose may not be available. Therefore, apart from the 3D geometric approach through gaze-ray–scene intersection, the PoG is commonly obtained through a mapping from the 3D eye pose, or eye or appearance features in the image. Early techniques require the user to perform a tedious per-measurement calibration of a mapping from image features to the PoG on a planar surface (Merchant, Morrissette, & Porterfield, 1974; Stampe, 1993). The strategy does not support for head movement and, thus, requires rigid body attachments, head motion compensation, or recalibration. Recent appearance-based techniques overcome the calibration requirement, and thus head-pose dependency, by introducing constraints such as gaze–mouse-cursor relation (Sugano, Matsushita, Sato, & Koike, 2008), and visual saliency (Sugano, Matsushita, & Sato, 2013), or applying a user-independent deep neural network, automatically trained in advance from a large dataset of photorealistically rendered synthetic images (Wood et al., 2015). Nevertheless, these methods require specific task and scene knowledge.

In order to obtain the PoG in the scene, 3D geometric eye-modeling-based systems suffer from two major drawbacks (Donegan et al., 2005; Hansen & Ji, 2010). First, mapping the pose of the eye to the PoG requires a manual *setup calibration* of the eye-camera–scene relation that suffers from potential user error and does not support dynamic setups. Second, due to the concept of PoG calculation by either intersecting the gaze ray with a static scene model or applying a calibrated mapping, systems suffer from a *parallax error* under depth-varying conditions (such as with complex scene geometry, dynamic, or mobile setups). This explains the common restriction to static setups and planar target scenes like monitors and walls, and imposes a difficulty for practical interactive application scenarios.

To overcome these issues, scene information can be directly obtained from corneal reflections. In case of planar scenes such as displays or projection screens, the reflection is identified or reconstructed from attached LEDs (Kang, Eizenman, Guestrin, 2008) or from the content itself (Hansen, Agustin, & Villanueva, 2010; Nitschke et al., 2011a; Schnieders et al., 2010). In case of arbitrary scenes, the PoG can be directly estimated in a corneal image, using either catadioptric back projection or forward projection. Both methods calculate the PoG or the complete subjective view from the corneal image or environment map. The back-projection method (Nishino & Nayar, 2006) projects the whole corneal image onto the retina, and extracts the PoG or subjective view from the retinal image. The forward-projection method (El Hafi et al., 2016; Nakazawa & Nitschke, 2012; Nitschke, 2011; Nitschke et al., 2013a; Takemura et al., 2014a, b) works in the opposite way, as it calculates where light around the PoG in the environment map reflects at the corneal surface into the eye image (Figure 21.15). This achieves increased accuracy as it does not depend on modeling of the inner eye, and allows for fast calculation of the PoG through single point projection rather than complete corneal image projection. However, it requires known scene distance or a compensation strategy.

The resulting PoG estimation approaches are based on two main ideas: First, the concept of the gaze-reflection point (GRP) describes where light from the PoG in the scene reflects at the corneal surface into an eye-observing image. The calculation applies the analytic solution for the forward-projection problem (Nakazawa & Nitschke, 2012; Nitschke, 2011; Nitschke et al., 2013a), which provides different strategies to cope with a commonly unknown scene distance. The model includes an optional one-time calibration and compensation for the individual offset between optical and visual axis. Takemura et al. (2014a, b) follow a different approach. They extract the corneal image as the tangent-plane projection at the optical axis (Takemura et al., 2014b) or visual axis, obtained through a one-time calibration (Takemura et al., 2014a). The GRP is calculated using standard analytic forward projection (Eberly, 2008), requiring known distance to a gazed planar object. When the distance is unknown, the GRP is approximated as the intersection of the optical or visual axis with the cornea, which commonly results in a larger gaze error compared to Nitschke (2011); Nakazawa and Nitschke (2012); Nitschke et al. (2013a). The concept of the GRP allows calculating the reflection of the PoG directly in an eye image, which may often be sufficient for visual analysis by a human expert.

Second, high-quality visual information of the gazed scene or computational analysis using generic techniques requires further processing to map information from the corneal image into a regular scene image or model. Takemura et al. (2014a, b) describe gazed object recognition for planar objects of same shape and size. Takemura et al. (2014b) use a database of object images, where a homography mapping is specified through the four manually selected corner points in the reprojected corneal image. Takemura et al. (2014a) improve on this to obtain an online strategy for gaze mapping and object recognition in a scene camera image, where the mapping is automatically obtained through template matching. For both methods, they show that the object that is gazed at can be successfully recognized through SIFT (Lowe, 2004) feature correspondence matching between the warped corneal image and the object or scene camera images. Nevertheless, the approaches suffer from two issues in practice: they do not offer robust handling of iris pattern noise and nonideal initial registration, and the registration strategy through dense template matching is computationally expensive. To overcome these issues, two approaches are used for robust image-based matching between the scene reflection in an eye image and a scene image or model. The two approaches target different application scenarios: the

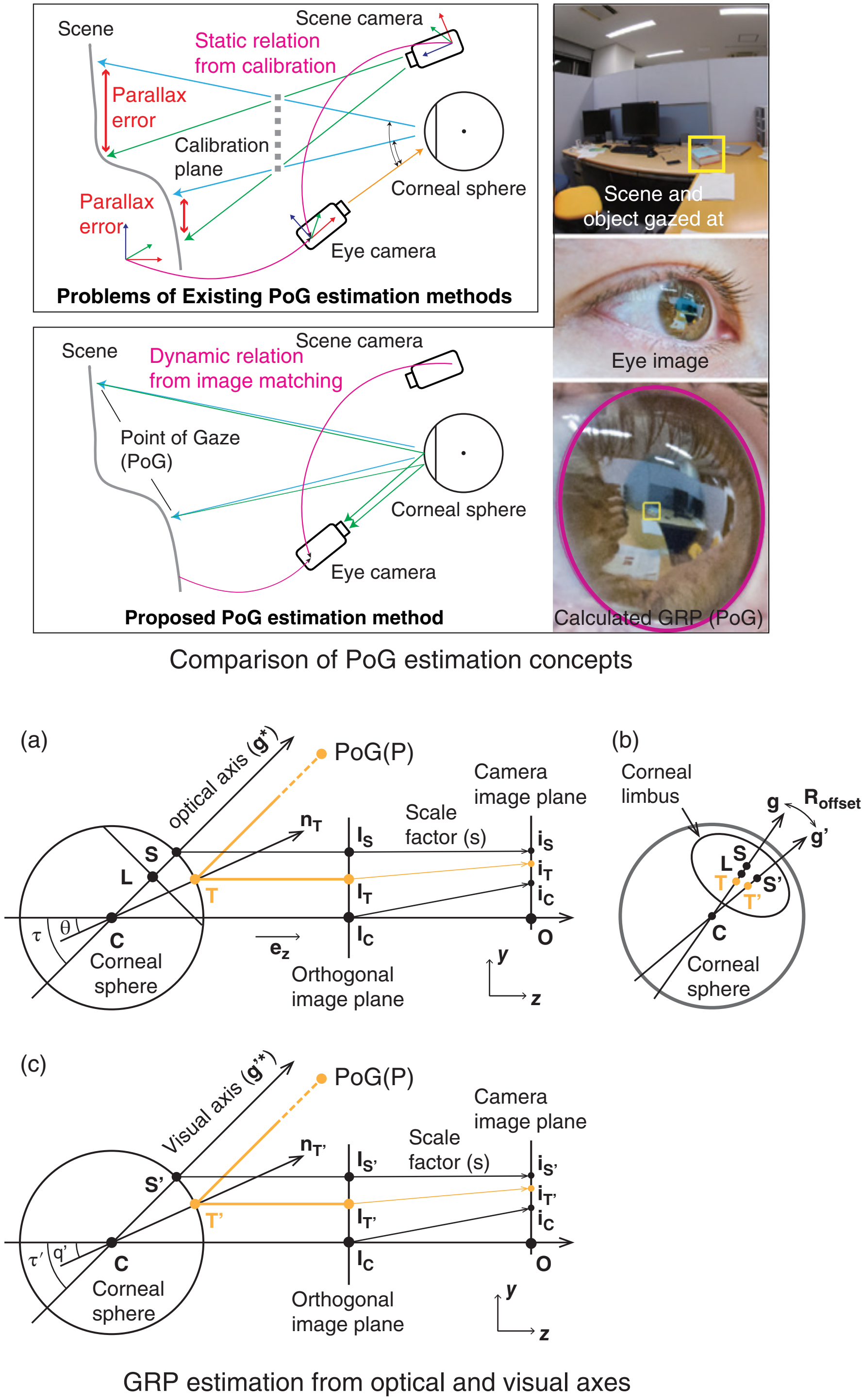


Figure 21.15 Nonintrusive point-of-gaze (PoG) tracking in arbitrary environments. (top box) Comparison of concepts: Conventional methods require a geometric calibration that causes parallax errors for a moving user and a nonplanar scene. The corneal-imaging-based method solves these problems, as it calculates the gaze reflection point (GRP) (PoG reflection) directly in an eye image and maps it into a scene image. The images show the concept of the GRP for a person looking

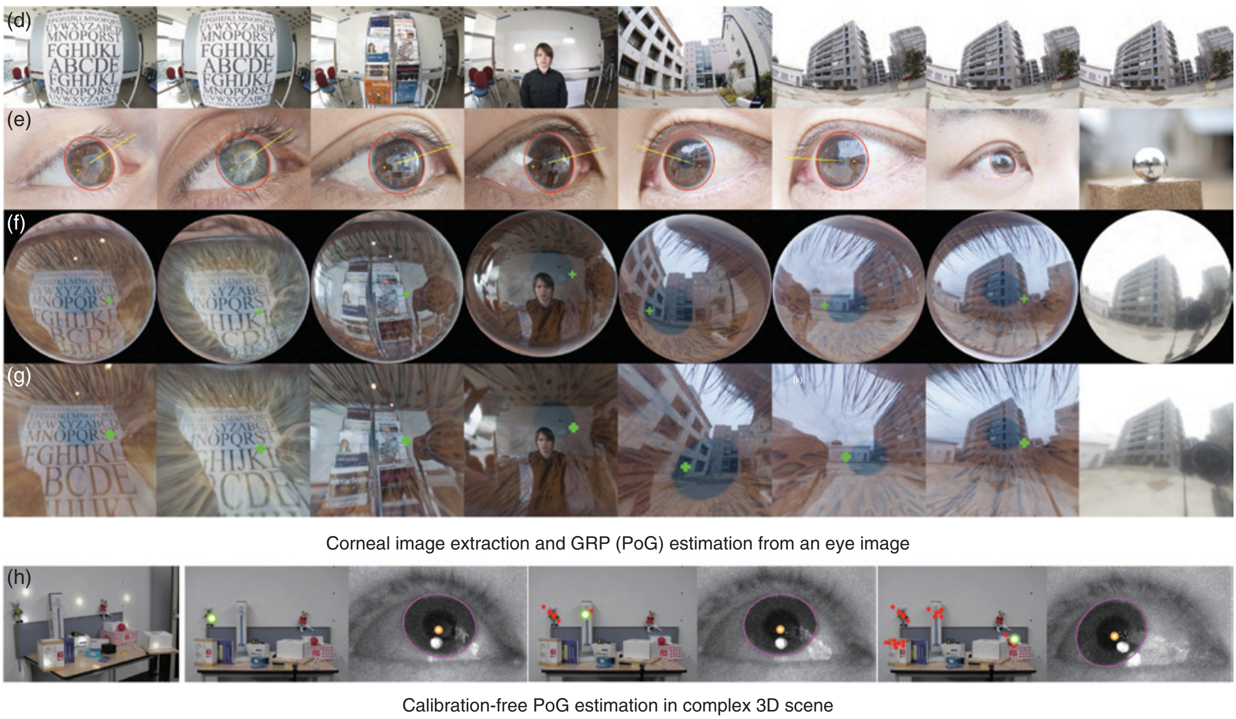


Figure 21.15 (Continued) at a book. (box, right) GRP estimation: The light ray towards the PoG and parallel to the gaze direction reflects at the GRP into an eye camera. (a) Assuming the gaze direction as the optical axis $\mathbf{g}$ results in the GRP $\mathbf{T}$. (b),(c) Calibrating the offset to the visual axis $\mathbf{g}'$ results in the corrected GRP $\mathbf{T}'$. (d)–(g) Corneal image extraction and PoG estimation: (d) Scene images. (e) (Cropped) eye images, with iris contour, gaze direction, and GRP. The right column shows a spherical mirror at the size of the cornea. (f) Spherical panorama (180°) and (g) reprojected perspective image (48°) with GRP. The quality of corneal reflections increases towards the apex, allowing for rich scene details near the GRP. (h) Calibration-free PoG estimation in a complex 3D scene using coded illumination. Three gaze sequence frames under free head motion show estimated PoG in the scene image (rendered dot) and corresponding GRP in the eye image (rendered dot).

active-light approach uses either an off-the-shelf digital video projector (Nitschke, 2011), a special high-power IR LED-array projector (Nakazawa & Nitschke, 2012) or LED markers (Nakazawa et al., 2017; Smith et al., 2005) to illuminate individual scene locations with a unique time series of coded illumination patterns. Decoding the illumination patterns from a scene image and the scene reflection in an eye image then obtains robust, accurate, and pixel-precise correspondence matches between eye image and scene image or markers. The passive approach (Nakazawa et al., 2016) does not require complex illumination hardware and works with any eye and scene imagery (that may be captured with a stationary camera and even at different times, such as with Google Street View). Robust correspondence matching is achieved by reducing the complexity of the problem through geometric modeling and random sampling. While this also reduces accuracy and precision, the effect is negligible for common practical usage scenarios. Moreover, the absence of illumination allows texture extraction from scene images and enables outdoor application under high brightness. The reduced hardware requirement allows compact setups, ideal for wearable cameras and augmented reality head-mounted displays.

In conclusion, corneal imaging methods enable ad hoc and nonintrusive EGT, with remote and head-mounted systems, in unconstrained dynamic environments and mobile scenarios. This flexibility opens up the potential for pervasive and everyday applications in the field and extends the possibilities to information retrieval from existing footage, such as in surveillance and forensics. This could be a powerful enabler for upcoming HCI solutions.

Future Directions

In this chapter, we have shown that geometric eye and corneal reflection modeling has the potential to facilitate novel technologies and application scenarios. While the field benefits from general achievements in catadioptric imaging (Sturm et al., 2011), practical realizations commonly demand tailored algorithms to cope with the unique properties of the human body. As a step in this direction, we recently observed a major shift from proof-of-concept to problem-oriented works that introduce more practical implementations. Nevertheless, there still remains a requirement for automatic, robust, and accurate solutions that can be applied in the field. Let us now discuss promising future directions, covering problems, methods, and applications. As these ideas relate to different stages of the processing pipeline, they may as well be combined.

Image acquisition

Capturing high-quality corneal reflections is a challenging task that requires expert knowledge or specific algorithms for parameter-adjustment (Nitschke & Nakazawa, 2012). Since the imaging setup is commonly time consuming, many potential application scenarios are likely to be missed. A promising solution is to exploit novel imaging technologies or combinations of hardware to do parallel recording and real-time parameter-adjustment, or to capture raw imagery with additional information and defer the adjustment towards postprocessing.

Regarding the first approach, pan-tilt-zoom (PTZ) cameras (Reale et al., 2010; Yoo & Chung, 2005), high-resolution camera arrays (Chong et al., 2017) and camera networks help improving resolution, field of view, and coverage. Chong et al. (2017) show an interesting setup that is successfully applied with infants in a long-term project to study behavioral patterns in autism spectrum disorders (ASD). To compensate for the inherent tradeoffs in corneal image capture, they engineered a sophisticated system that integrates a wide field-of-view RGB-D camera with a 2×3 array of high-resolution narrow field-of-view and depth-of-field cameras. The RGB-D camera performs face tracking and provides the camera array with eye-pose information for real-time focus adjustment and region-of-interest calculation. While the system achieves impressive results, its complexity and cost impose severe limitations.

Regarding the second approach, computational photography achieves flexibility to tradeoff and overcome shortcomings in conventional photography through combining optical and computational techniques: Light-field (LF) or plenoptic cameras (Lytro, Lytro Light Field imaging platform, https://www.lytro.com; Raytrix, Raytrix Light Field Technology, https://raytrix.de) trade spatial resolution for additional directional information of light rays, which allows removing defocus blur and improving eye and face pose estimation. High-dynamic-range (HDR) imaging techniques (Debevec & Malik, 1997) and cameras (El Hafi et al., 2016; ViewPLUS, ViewPLUS XViii 18bit camera, http://www.viewplus.co.jp) trade temporal resolution for sensitivity, which allows the improvement of contrast, noise, and defocus blur. With LF and HDR imaging, the ability for digital refocusing and aperture adjustment tremendously simplifies and even enables corneal image acquisition for scenarios where conventional techniques would fail. While first realizations are promising, for example to reveal features under low-light conditions with HDR techniques (El Hafi et al., 2016), it will take time to see practical implementations, as the current state of the art still introduces great complexity regarding development, setup, and data processing.

Scene information

Scene information can provide additional knowledge for robustly solving geometric modeling problems in corneal imaging. Assigning constraints to tracked scene features enables the recovery of unknown information and increases the performance of algorithms. Geometric constraints from reflected 3D points, lines and planes that commonly occur in man-made environments are well suited. Standard PCCR methods in eye-gaze tracking (EGT) apply point features (glints) from known light sources for eye-pose estimation (Guestrin & Eizenman, 2006; Shih et al., 2000). The approach may be extended to the illumination from an unknown scene where structure information could be obtained using stereo or structure-from-motion (SfM) approaches. A 3D line, imaged by a catadioptric system, allows recovering the line, as well as the pose and shape of an axisymmetric mirror (Gasparini & Caglioti, 2011; Lanman, Wachs, Taubin, & Cukierman, 2006b). Combining constraints from lines and planes enables recovering planar polygonal light sources (Schnieders, Wong, & Dai, 2009). Line and plane constraints are less common in corneal reflection analysis, and have been exploited only for display pose estimation and EGT (Nitschke et al., 2011a; Schnieders et al., 2010). Nevertheless, there exists a wide range of related literature on catadioptric imaging, regarding the recovery of mirror pose, shape, and scene geometry (Sturm et al., 2011).

Finally, all measurements may be integrated into a bundle-adjustment framework to estimate and optimize unknown parameters, probably in conjunction with outlier removal.

Eye information

Eye pose and shape estimation using scene correspondence matches and constraints have been largely discussed. If an environment map of scene illumination is available, a different strategy, similar to Chang, Raskar, and Agrawal (2009) may be applied, where an actual eye image is matched against a database of synthetic specular highlight or flow images, rendered offline for the complete search space of eye poses and shapes. As with scene information, the modeling and registration of body information (such as face and eye features) is important. Eye features are anatomic point, contour or texture features of the eyeball or the occluding contour of the skin, related to pupil, iris, sclera, eye lids or eye corners. Pupil and iris features also require handling of corneal refraction, which depends on the shape and pose of the cornea. Nearly all methods in geometric eye modeling assume the cornea to be spherical. However, as the eyeball is slightly flattened in the vertical plane (Snell & Lemp, 1997) and the corneal topology is complex (Bogan et al., 1990; Gatinel et al., 2011), it is necessary to develop strategies for parameterizing and calibrating more accurate models. Existing aspherical approaches assume an axisymmetric shape and require known scene positions (Nagamatsu et al., 2010). Further improvements may be achieved with more general parametric models that are calibrated using generic scene constraints (Gasparini & Caglioti, 2011) or image matching (Nakazawa et al., 2016). More detailed modeling of the light interaction at the eye, such as reflection at the retina and iris texture refraction at the cornea, may also improve modeling and tracking. Finally, an interesting direction in eye analysis is to develop solutions that combine low-cost optical hardware with personal devices to enable self-assessment of visual conditions, such as refractive errors, focal range, focusing speed and lens opacity (Pamplona et al., 2010, 2011). The results may also be applied to correct the visual content in digital devices.

Alternative approaches

This chapter mainly focused on the computational analysis of corneal reflection imagery using specific algorithms for geometric modeling, image processing, and information extraction. It shows that automated realtime systems can facilitate a wide range of applications. However, corneal image processing is a challenging and specialized task, where fully automated techniques are not yet available and only slowly progressing. In order to realize practical solutions at the moment, it is necessary to investigate flexible application-specific approaches such as those involving problem simplification and transformation, generic computational methods or human-in-the-loop, to model and mine information from corneal reflections. A promising strategy could be to identify how robust generic computer vision methods can be applied with imperfect imagery such as raw or partially processed corneal reflections. Such an approach has been successfully applied by several groups. In the context of information security, Backes et al. (2008, 2009) show the feasibility of eavesdropping by recovering information from corneal reflections of eye images captured at a distance. Though, operating directly on the eye images, they successfully compensate for defocus and motion blur using standard nonblind deconvolution methods. Regarding the ongoing interest and progress in wearable technology, Lander et al. (2016) recently built a head-mounted eye-imaging system to capture a two-day data series in a lifelogging scenario. Semiautomated postprocessing of the raw eye imagery using standard techniques allows the analysis of human-to-human interaction (through face detection), display interaction (through modified square detection), and interaction with generic objects having characteristics similar to displays, such as posters, books, and shelves (through object detection by template matching).

From a broader perspective, corneal reflections do not only have application potential but also the potential to provide solutions to pressing current problems. It is therefore necessary to focus on the practicality of techniques and implementations. In fact, reviewing the development of the field, we find that the scope of works evolves towards more application-oriented settings. While it is important to continue improving eye image processing and modeling, this should not be done just for the sake of it. Rather, it is beneficial to analyze a corneal-imaging problem starting from its application. Determining requirements for the result and choosing the solution as a tradeoff between accuracy and practicality allows generic techniques to be considered. Following this approach, an implementation could benefit from general technical advancement happening at high pace, such as with ambient/wearable sensors and cameras (size, resolution, positioning, zoom, focus, aperture), and processing algorithms (especially considering "big data" and machine learning). Following such a strategy could allow development, setup and system complexity to be reduced—and in many cases may be necessary to exploit information from corneal reflections at all.

Further broadening the perspective, information from corneal reflections could be also beneficial for use cases apart from a computational analysis. For example, as human perception is very tolerant to poor image quality, Jenkins and Kerr (2013) show that this allows information recognition from low-quality corneal reflections such as face recognition for identity discrimination, familiar face recognition, and even emotional state and interest recognition tasks. They also made the observation that optimal viewers (who are familiar with the faces to be recognized) may be more important than optimal imagery. The general conclusion could be that chasing for optimal (computational or human) recognizers can be more promising than for optimal (corneal) image acquisition and processing.

Conclusion

This chapter provided a comprehensive overview on the modeling, processing and analysis of corneal reflections from environmental illumination with the aim of raising awareness of the topic and encouraging novel development. The contents covered the entire processing pipeline, comprising an introduction to the anatomic background, a technical description of geometric eye modeling, corneal reflection modeling, and high-level corneal image processing, an overview of applications, and a discussion of future directions.[7]

[7] The interested reader may further refer to Nitschke (2011) for a more detailed coverage of the topic.

We discussed relevance to a number of fields, including visual recognition, computer graphics and vision, HCI, and diagnostic studies. Advancement in computational systems, devices, and architectures demands novel interfaces and forms of interaction. Corneal reflection analysis relates the individual (eye) with the physical world, which facilitates a flexible framework for tracking cameras, eyes, and visual scene information to solve various problems. This inherently allows for the ad hoc and dynamic application of interaction techniques in the field.

While corneal imaging and applications are promising, practical implementations still require more capable models and solutions to technical issues for achieving robust processing of unconstrained eye and scene information. Regarding this, we discussed relevant future directions comprising improved hardware, problem formulation, alternative approaches, and interesting applications.

References

Agrawal, A., Taguchi, Y., & Ramalingam, S. (2011). Beyond Alhazen's problem: Analytical projection model for non-central catadioptric cameras with quadric mirrors. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 2993–3000). doi:10.1109/CVPR.2011.5995596

Agustin, J. S., Villanueva, A., & Cabeza, R. (2006). Pupil brightness variation as a function of gaze direction. In *Proceedings of the ACM Symposium on Eye Tracking Research and Applications (ETRA)* (p. 49). doi:10.1145/1117309.1117334

Arican, Z., & Frossard, P. (2011). Joint registration and super-resolution with omnidirectional images. *IEEE Transactions on Image Processing, 20*(11), 3151–3162.

Arvacheh, E. M., & Tizhoosh, H. R. (2006). IRIS segmentation: Detecting pupil, limbus and eyelids. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)* (pp. 2453–2456). doi:10.1109/ICIP.2006.312773

Atchison, D. A., & Smith, G. (2000). *Optics of the human eye*. Edinburgh: Butterworth-Heinemann.

Axholt, M., Skoglund, M. A., O'Connell, S. D., Cooper, M. D., Ellis, S. R., & Ynnerman, A. (2011). Parameter estimation variance of the single point active alignment method in optical see-through head mounted display calibration. In *Proceedings of the IEEE Virtual Reality Conference (VR)* (pp. 27–34). doi:10.1109/VR.2011.5759432

Backes, M., Chen, T., Dürmuth, M., Lensch, H. P. A., & Welk, M. (2009). Tempest in a teapot: Compromising reflections revisited. In *Proceedings of the IEEE Symposium on Security and Privacy (SP)* (pp. 315–327). doi:10.1109/SP.2009.20

Backes, M., Dürmuth, M., & Unruh, D. (2008). Compromising reflections-or-how to read LCD monitors around the corner. In *Proceedings of the IEEE Symposium on Security and Privacy (SP)* (pp. 158–169). doi:10.1109/SP.2008.25

Bakaraju, R. C., Ehrmann, K., Papas, E., & Ho, A. (2008). Finite schematic eye models and their accuracy to in-vivo data. *Vision Research, 48*(16), 1681–1694.

Baker, S., & Nayar, S. K. (1999). A theory of single-viewpoint catadioptric image formation. *International Journal of Computer Vision, 35*(2), 175–196.

Baker, T. Y. (1943). Ray tracing through non-spherical surfaces. *Proceedings of the Physical Society, 55*(5), 361–364.

Balzer, J., & Werling, S. (2010). Principles of shape from specular reflection. *Measurement, 43*(10), 1305–1317.

Barnard, K., Cardei, V., & Funt, B. (2002). A comparison of computational color constancy algorithms. I: Methodology and experiments with synthesized data. *IEEE Transactions on Image Processing, 11*(9), 972–984.

Barnard, K., Martin, L., Coath, A., & Funt, B. (2002). A comparison of computational color constancy Algorithms. II. Experiments with image data. *IEEE Transactions on Image Processing, 11*(9), 985–996.

Barry, J. C., Pongs, U. M., & Hillen, W. (1997). Algorithm for purkinje images I and IV and limbus centre localization. *Computers in Biology and Medicine, 27*(6), 515–531.

Bay, H., Tuytelaars, T., & Gool, L. V. (2006). SURF: speeded up robust features. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 404–417). doi:10.1007/11744023_32

Bérard, P., Bradley, D., Nitti, M., Beeler, T., & Gross, M. (2014). High-quality capture of eyes. *ACM Transactions on Graphics, 33*(6), 223:1–223:12.

Beymer, D., & Flickner, M. (2003). Eye gaze tracking using an active stereo head. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 451–458). doi:10.1109/CVPR.2003.1211502

Bogan, S. J., Waring III, G. O., Ibrahim, O., Drews, C., & Curtis, L. (1990). Classification of normal corneal topography based on computer-assisted videok eratography. *Archives of Ophthalmology, 108*(7), 945–949.

Bolt, R. A. (1982). Eyes at the interface. In *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI)* (pp. 360–362). doi:10.1145/800049.801811

Bonfort, T., Sturm, P., & Gargallo, P. (2006). General specular surface triangulation. In *Proceedings of the Asian Conference on Computer Vision (ACCV)* (pp. 872–881). doi:10.1007/11612704_87

Bowyer, K. W., Hollingsworth, K., & Flynn, P. J. (2008). Image understanding for iris biometrics: A survey. *Computer Vision and Image Understanding, 110*(2), 281–307.

Bruckstein, A. M., & Richardson, T. J. (2000). Omniview cameras with curved surface mirrors. In *Proceedings of the IEEE Workshop on Omnidirectional Vision (OMNIVIS)* (pp. 79–84). doi:10.1109/OMNVIS.2000.853809

Buchsbaum, G. (1980). A spatial processor model for object colour perception. *Journal of the Franklin Institute, 310*(1), 1–26.

Bulling, A., Roggen, D., & Tröster, G. (2009). Wearable EOG goggles: Seamless sensing and context-awareness in everyday environments. *Journal of Ambient Intelligence and Smart Environments, 1*(2), 157–171.

Caglioti, V., Taddei, P., Boracchi, G., Gasparini, S., & Giusti, A. (2007). Single-image calibration of off-axis catadioptric cameras using lines. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 1–6). doi:10.1109/ICCV.2007.4409192

Caudell, T., & Mizell, D. (1992). Augmented reality: An application of heads-up display technology to manual manufacturing processes. In *Proceedings of the Hawaii International Conference on System Sciences* (pp. 659–669). doi:10.1109/HICSS.1992.183317

Chang, J. Y., Raskar, R., & Agrawal, A. (2009). 3D pose estimation and segmentation using specular cues. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1706–1713). doi:10.1109/CVPR.2009.5206820

Chen, J., & Ji, Q. (2008). 3D gaze estimation with a single camera without IR illumination. In *Proceedings of the IEEE International Conference on Pattern Recognition (ICPR)* (pp. 1–4). doi:10.1109/ICPR.2008.4761343

Chen, Q., Wu, H., & Wada, T. (2004). Camera calibration with two arbitrary coplanar circles. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 521–532). doi :10.1007/978-3-540-24672-5_41

Chong, E., Nitschke, C., Nakazawa, A., Rozga, A., & Rehg, J. M. (2017). *Noninvasive corneal image-based gaze measurement system*. Retrieved from https://arxiv.org/abs/1708.00908v1

Clark, J. J. (2010). Photometric stereo using LCD displays. *Image and Vision Computing, 28*(4), 704–714.

Colombo, C., Comanducci, D., & Bimbo, A. D. (2007). Robust tracking and remapping of eye appearance with passive computer vision. *ACM Transactions on Multimedia Computing, Communications, and Applications, 3*(4), 2:1–2:20.

Crick, R. P., & Khaw, P. T. (2003). *A textbook of clinical ophthalmology* (3rd ed.). Singapore: World Scientific.

Daugman, J. (2004). How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology, 14*(1), 21–30.

Debevec, P. E., & Malik, J. (1997). Recovering high dynamic range radiance maps from photographs. In *Proceedings of the ACM SIGGRAPH '97* (pp. 369–378). doi:10.1145/1401132.1401174

Donegan, M., Oosthuizen, L., Bates, R., Daunys, G., Hansen, J., Joos, M.,… Signorile, I. (2005). *Cogain: D3.1 user requirements report, with observations of difficulties users are experiencing*. (Technical report, European Union Network of Excellence COGAIN, contract no. IST-2003-511598, of the 6th Framework Programme). Retrieved from http://wiki.cogain.org/images/e/ef/COGAIN-D3.1.pdf

Duchowski, A. T. (2002). A breadth-first survey of eye-tracking applications. *Behavior Research Methods, Instruments, and Computers, 34*(4), 455–470.

Duchowski, A. T. (2007). *Eye Tracking Methodology: Theory and Practice* (2nd ed.). London: Springer-Verlag.

Eberly, D. (2008). *Computing a point of reflection on a sphere*. Retrieved from http://www.geometrictools.com/Documentation/SphereReflections.pdf

Ebisawa, Y. (1998). Improved video-based eye-gaze detection method. *IEEE Transactions on Instrumentation and Measurement, 47*(4), 948–955.

Einighammer, J., Oltrup, T., Bende, T., & Jean, B. (2009). The individual virtual eye: A computer model for advanced intraocular lens calculation. *Journal of Optometry, 2*(2), 70–82.

El Hafi, L., Ding, M., Takamatsu, J., & Ogasawara, T. (2016). Gaze tracking using corneal images captured by a single high-sensitivity camera. In *Proceedings of the IET IBC Conference* (pp. 33(10.)–33(10.)(1)). doi:10.1049/ibc.2016.0033

Escudero-Sanz, I. & Navarro, R. (1999). Off-axis aberrations of a wide-angle schematic eye model. *Journal of the Optical Society of America, 16*(8), 1881–1891.

Farsiu, S., Robinson, M. D., Elad, M., & Milanfar, P. (2004). Fast and robust multiframe super resolution. *IEEE Transactions on Image Processing, 13*(10), 1327–1344.

Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM, 24*(6), 381–395.

Fitzgibbon, A., Pilu, M., & Fisher, R. B. (1999). Direct least square fitting of ellipses. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 21*(5), 476–480.

Fofi, D., Sliwa, T., & Voisin, Y. (2004). A comparative survey on invisible structured light. In *Proceedings of the SPIE 5303 Machine Vision Applications in Industrial Inspection XII* (pp. 90–98). doi:10.1117/12.525369

Francken, Y., Cuypers, T., Mertens, T., Gielis, J., & Bekaert, P. (2008). High quality mesostructure acquisition using specularities. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1–7). doi:10.1109/CVPR.2008.4587782

Francken, Y., Hermans, C., & Bekaert, P. (2009). Screen-camera calibration using Gray codes. In *Proceedings of the IEEE Canadian Conference on Computer and Robot Vision (CRV)* (pp. 155–161). doi:10.1109/CRV.2009.35

Francken, Y., Hermans, C., Cuypers, T., & Bekaert, P. (2008). Fast normal map acquisition using an LCD screen emitting gradient patterns. In *Proceedings of the IEEE Canadian Conference on Computer and Robot Vision (CRV)* (pp. 189–195). doi:10.1109/CRV.2008.22

Francois, G., Gautron, P., Breton, G., & Bouatouch, K. (2009). Image-based modeling of the human eye. *IEEE Transactions on Visualization and Computer Graphics, 15*(5), 815–827.

Funk, N., & Yang, Y.-H. (2007). Using a raster display for photometric stereo. In *Proceedings of the IEEE Canadian Conference on Computer and Robot Vision (CRV)* (pp. 201–207). doi:10.1109/CRV.2007.66

Gasparini, S., & Caglioti, V. (2011). Line localization from single catadioptric images. *International Journal of Computer Vision, 94*(3), 361–374.

Gatinel, D., Malet, J., Hoang-Xuan, T., & Azar, D. T. (2011). Corneal elevation topography: Best fit sphere, elevation distance, asphericity, toricity, and clinical implications. *Cornea, 30*(5), 508–515.

Geyer, C., & Daniilidis, K. (2002). Properties of the catadioptric fundamental matrix. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 536–538). doi:10.1007/3-540-47967-8_10

Google, Google Street View – Explore natural wonders and world landmarks https://www.google.com/streetview/

Goncalves, N., & Araujo, H. (2004). Projection model, 3D reconstruction and rigid motion estimation from non-central catadioptric images. In *Proceedings of the International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT)* (pp. 325–332). doi:10.1109/TDPVT.2004.1335221

Grabowski, K., Sankowski, W., Zubert, M., & Napieralska, M. (2006). Reliable iris localization method with application to iris recognition in near infrared light. In *Proceedings of the IEEE International Conference on Mixed Design of Integrated Circuits and System (MIXDES)* (pp. 684–687). doi:10.1109/MIXDES.2006.1706671

Guestrin, E. D., & Eizenman, M. (2006). General theory of remote gaze estimation using the pupil center and corneal reflections. *IEEE Transactions on Biomedical Engineering, 53*(6), 1124–1133.

Guestrin, E. D., & Eizenman, M. (2011). Remote point-of-gaze estimation with single-point personal calibration based on the pupil boundary and corneal reflections. In *Proceedings of the IEEE Canadian Conference on Electrical and Computer Engineering (CCECE)* (pp. 000971–000976). doi:10.1109/CCECE.2011.6030604

Guillon, M., Lydon, D. P., & Wilson, C. (1986). Corneal topography: A clinical model. *Ophthalmic and Physiological Optics, 6*(1), 47–56.

Gullstrand, A. (1909). *Helmholtz's Handbuch der Physiologischen Optik*. Hamburg: Voss.

Gurdjos, P., Sturm, P., & Wu, Y. (2006). Euclidean structure from $N \geq 2$ parallel circles: Theory and algorithms. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 238–252). doi:10.1007/11744023_19

Halir, R., & Flusser, J. (1998). Numerically stable direct least squares fitting of ellipses. In *Proceedings of the International Conference in Central Europe on Computer Graphics and Visualization (WSCG)* (pp. 125–132). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.1.7559

Halstead, M. A., Barsky, B. A., Klein, S. A., & Mandell, R. B. (1996). Reconstructing curved surfaces from specular reflection patterns using spline surface fitting of normals. In *Proceedings of the ACM SIGGRAPH '96* (pp. 335–342). doi:10.1145/237170.237272

Hansen, D. W., Agustin, J. S., & Villanueva, A. (2010). Homography normalization for robust gaze estimation in uncalibrated setups. In *Proceedings of the ACM Symposium on Eye Tracking Research and Applications (ETRA)* (pp. 13–20). doi:10.1145/1743666.1743670

Hansen, D. W., & Ji, Q. (2010). In the eye of the beholder: A survey of models for eyes and gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 32*(3), 478–500.

Hartley, R., & Zisserman, A. (2003). *Multiple View Geometry in Computer Vision* (2nd ed.). New York, NY: Cambridge University Press.

He, Z., Tan, T., Sun, Z., & Qiu, X. (2009). Toward accurate and fast iris segmentation for iris biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 31*(9), 1670–1684.

Hennessey, C., Noureddin, B., & Lawrence, P. (2006). A single camera eye-gaze tracking system with free head motion. In *Proceedings of the ACM Symposium on Eye Tracking Research and Applications (ETRA)* (pp. 87–94). doi:10.1145/1117309.1117349

Holloway, R. L. (1997). Registration error analysis for augmented reality. *Presence: Teleoperators and Virtual Environments, 6*(4), 413–432.

Holmberg, N. (2012). *Advance head rig.* Retrieved from http://www.blendswap.com/blends/view/48717.

Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Halszka, J., & van de Weijer, J. (2011). *Eye Tracking: A comprehensive guide to methods and measures.* Oxford: Oxford University Press.

Hua, H., Krishnaswamy, P., & Rolland, J. P. (2006). Video-based eyetracking methods and algorithms in head-mounted displays. *Optics Express, 14*(10), 4328–4350.

Ihrke, I., Kutulakos, K. N., Lensch, H. P. A., Magnor, M. A., & Heidrich, W. (2010). Transparent and specular object reconstruction. *Computer Graphics Forum, 29*(8), 2400–2426.

Isard, M., & Blake, A. (1998). Condensation–conditional density propagation for visual tracking. *International Journal of Computer Vision, 29*(1), 5–28.

Ishiguro, Y., Mujibiya, A., Miyaki, T., & Rekimoto, J. (2010). Aided eyes: Eye activity sensing for daily life. In *Proceedings of the Augmented Human International Conference (AH)* (pp. 25:1–25:7). doi:10.1145/1785455.1785480

Iskander, D. (2006). A parametric approach to measuring limbus corneae from digital images. *IEEE Transactions on Biomedical Engineering, 53*(6), 1134–1140.

Iskander, D. R., Collins, M. J., Mioschek, S., & Trunk, M. (2004). Automatic pupillometry from digital images. *IEEE Transactions on Biomedical Engineering, 51*(9), 1619–1627.

Itoh, Y. & Klinker, G. (2014). Interaction-free calibration for optical see-through head-mounted displays based on 3d eye localization. In *Proceedings of the IEEE Symposium on 3D User Interfaces (3DUI)* (pp. 75–82). doi:10.1109/3DUI.2014.6798846

Jaimes, A. & Sebe, N. (2007). Multimodal human–computer interaction: A survey. *Computer Vision and Image Understanding, 108*(1–2), 116–134.

Jenkins, R., & Kerr, C. (2013). Identifiable images of bystanders extracted from corneal reflections. *PLOS ONE, 8*(12), 1–5.

Jimenez, J., Jarabo, A., Gutierrez, D., Danvoye, E., & Pahlen, J. v. d. (2012). Separable subsurface scattering and photorealistic eyes rendering. In *Proceedings of the ACM SIGGRAPH '12 Courses: Advances in Real-Time Rendering* [Course notes.]. New York, NY: ACM.

Johnson, M. K., & Farid, H. (2007). Exposing digital forgeries through specular highlights on the eye. In *Proceedings of the International Workshop on Information Hiding (IH)* (pp. 311–325). doi:10.1007/978-3-540-77370-2_21

Kaehler, A., & Bradski, G. (2015). *Learning OpenCV 3: Computer vision in C++ with the OpenCV Library.* Sebastopol, CA: O'Reilly Media.

Kagami, S. (2010). High-speed vision systems and projectors for real-time perception of the world. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 100–107). doi:10.1109/CVPRW.2010.5543776

Kanatani, K., & Liu, W. (1993). 3D interpretation of conics and orthogonality. *CVGIP: Image Understanding*, *58*(3), 286–301.

Kang, J. J., Eizenman, M., Guestrin, E. D., & Eizenman, E. (2008). Investigation of the cross-ratios method for point-of-gaze estimation. *IEEE Transactions on Biomedical Engineering, 55*(9), 2293–2302.

Kaufman, P. L. & Alm, A. (2003). *Adler's physiology of the eye: Clinical application* (10th ed.). St. Louis, MO: Mosby.

Khurana, A. K. (2007). *Comprehensive ophthalmology* (4th ed.). New Delhi: New Age International (P) Ltd.

Kim, S. M., Sked, M., & Ji, Q. (2004). Non-intrusive eye gaze tracking under natural head movements. In *Proceedings of the International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS)* (pp. 2271–2274). doi:10.1109/IEMBS.2004.1403660

Kishishita, N., Kiyokawa, K., Orlosky, J., Mashita, T., Takemura, H., & Kruijff, E. (2014). Analysing the effects of a wide field of view augmented reality display on search performance in divided attention tasks. In *Proceedings of the IEEE/ACM International*

Symposium on Mixed and Augmented Reality (ISMAR) (pp. 177–186). doi:10.1109/ISMAR.2014.6948425

Kohlbecher, S., Bardinst, S., Bartl, K., Schneider, E., Poitschke, T., & Ablassmeier, M. (2008). Calibration-free eye tracking by reconstruction of the pupil ellipse in 3D space. In *Proceedings of the ACM Symposium on Eye Tracking Research and Applications (ETRA)* (pp. 135–138). doi:10.1145/1344471.1344506

Kooijman, A. C. (1983). Light distribution on the retina of a wide-angle theoretical eye. *Journal of the Optical Society of America, 73*(11), 1544–1550.

Kumar, R. K., Ilie, A., Frahm, J.-M., & Pollefeys, M. (2008). Simple calibration of non-overlapping cameras with a mirror. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1–7). doi:10.1109/CVPR.2008.4587676

Kutulakos, K. N., & Steger, E. (2008). A theory of refractive and specular 3D shape by light-path triangulation. *International Journal of Computer Vision, 76*(1), 13–29.

Lam, M. W. Y., & Baranoski, G. V. G. (2006). A predictive light transport model for the human iris. *Computer Graphics Forum, 25*(3), 359–368.

Lambooij, M., IJsselsteijn, W., Fortuin, M., & Heynderickx, I. (2009). Visual discomfort and visual fatigue of stereoscopic displays: A review. *Journal of Imaging Science and Technology, 53*(3), 1–14.

Lander, C., Krüger, A., & Löchtefeld, M. (2016). "The story of life is quicker than the blink of an eye": Using corneal imaging for life logging. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct (UbiComp)* (pp. 1686–1695). doi:10.1145/2968219.2968337

Lanman, D., Crispell, D., Wachs, M., & Taubin, G. (2006a). Spherical catadioptric arrays: Construction, multi-view geometry, and calibration. In *Proceedings of the International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT)* (pp. 81–88). doi:10.1109/3DPVT.2006.130

Lanman, D., Wachs, M., Taubin, G., & Cukierman, F. (2006b). Reconstructing a 3D line from a single catadioptric image. In *Proceedings of the International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT)* (pp. 89–96). doi:10.1109/3DPVT.2006.115

Lee, J. C., Hudson, S., & Dietz, P. (2007). Hybrid infrared and visible light projection for location tracking. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)* (pp. 57–60). doi:10.1145/1294211.1294222

Lefohn, A., Budge, B., Shirley, P., Caruso, R., & Reinhard, E. (2003). An ocularist's approach to human iris synthesis. *IEEE Computer Graphics and Applications, 23*(6), 70–75.

Le Grand, Y., & El Hage, S. G. (1980). *Physiological optics.* Berlin: Springer.

Lhuillier, M. (2008). Automatic scene structure and camera motion using a catadioptric system. *Computer Vision and Image Understanding, 109*(2), 186–203.

Li, D., Winfield, D., & Parkhurst, D. J. (2005). Starburst: A hybrid algorithm for video-based eye tracking combining feature-based and model-based approaches. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 79–86). doi:10.1109/CVPR.2005.531

Li, F., Kolakowski, S., & Pelz, J. (2007). Using structured illumination to enhance video-based eye tracking. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)* (pp. 373–376). doi:10.1109/ICIP.2007.4378969

Liestol, G. & Morrison, A. (2013). Views, alignment and incongruity in indirect augmented reality. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality—Arts, Media, and Humanities (ISMAR-AMH)* (pp. 23–28). doi:10.1109/ISMAR-AMH.2013.6671263

Lindsay, R., Smith, G., & Atchison, D. (1998). Descriptors of corneal shape. *Optometry and Vision Science, 75*(2), 156–158.

Liou, H. L. & Brennan, N. A. (1997). Anatomically accurate, finite model eye for optical modeling. *Journal of the Optical Society of America A, 14*(8), 1684–1695.

Lotmar, W. (1971). Theoretical eye model with aspherics. *Journal of the Optical Society of America, 61*(11), 1522–1529.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision, 60*(2), 91–110.

Machado, C. J., & Nelson, E. E. (2011). Eye-tracking with nonhuman primates is now more accessible than ever before. *American Journal of Primatology, 73*(6), 562–569.

Mandell, R. B. (1996). A guide to videokeratography. *International Contact Lens Clinic, 23*(6), 205–228.

Matas, J., Chum, O., Urban, M., & Pajdla, T. (2002). Robust wide baseline stereo from maximally stable extremal regions. In *Proceedings of the British Machine Vision Conference (BMVC)* (pp. 36.1–36.10). doi:10.5244/C.16.36

Matey, J. R., Broussard, R., & Kennell, L. (2010). Iris image segmentation and sub-optimal images. *Image and Vision Computing, 28*(2), 215–222.

Matsumoto, D., Keltner, D., Shiota, M. N., O'Sullivan, M., & Frank, M. (2008). Facial expressions of emotion. In M. Lewis, J. M. Haviland-Jones & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 211–234). New York, NY: Guilford.

Merchant, J., Morrissette, R., & Porterfield, J. L. (1974). Remote measurement of eye direction allowing subject motion over one cubic foot of space. *IEEE Transactions on Biomedical Engineering, 21*(4), 309–317.

Moon, T. K. (2005). Error correction coding: Mathematical methods and algorithms. Hoboken, NJ: Wiley-Interscience.

Morimoto, C. H., Koons, D., Amir, A., & Flickner, M. (2000). Pupil detection and tracking using multiple light sources. *Image and Vision Computing, 18*(4), 331–335.

Morimoto, C. H., & Mimica, M. R. M. (2005). Eye gaze tracking techniques for interactive applications. *Computer Vision and Image Understanding, 98*(1), 4–24.

Morris, N. J. W., & Kutulakos, K. N. (2007). Reconstructing the surface of inhomogeneous transparent scenes by scatter-trace photography. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 1–8). doi:10.1109/ICCV.2007.4408882

Moser, K., Itoh, Y., Oshima, K., Swan, J., Klinker, G., & Sandor, C. (2015). Subjective evaluation of a semi-automatic optical see-through head-mounted display calibration technique. *IEEE Transactions on Visualization and Computer Graphics, 21*(4), 491–500.

Motion Pictures. (2004). Producers: Sarafian, D., & Sinise, G. Director: Sarafian, D. CSI: NY: Season 1 Episode 10: Night, Mother [Motion picture.], USA, Paramount Pictures.

Nagahara, H., Yagi, Y., & Yachida, M. (2001). Resolution improving method from multi-focal omnidirectional images. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)* (pp. 654–657). doi:10.1109/ICIP.2001.959130

Nagamatsu, T., Iwamoto, Y., Kamahara, J., Tanaka, N., & Yamamoto, M. (2010). Gaze estimation method based on an aspherical model of the cornea: Surface of revolution about the optical axis of the eye. In *Proceedings of the ACM Symposium on Eye Tracking Research and Applications (ETRA)* (pp. 255–258). doi:10.1145/1743666.1743726

Nagamatsu, T., Kamahara, J., Iko, T., & Tanaka, N. (2008). One-point calibration gaze tracking based on eyeball kinematics using stereo cameras. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA)* (pp. 95–98). doi:10.1145/1344471.1344496

Nakazawa, A. (2016). Noise stable image registration using random resample consensus. In *Proceedings of the IEEE International Conference on Pattern Recognition (ICPR)* (pp. 1–6). doi:10.1109/ICPR.2016.7899742

Nakazawa, A., Kato, H., Nitschke, C., & Nishida, T. (2017). Eye gaze tracking using corneal imaging and active illumination devices. *Advanced Robotics, 31*(8), 413–427.

Nakazawa, A. & Nitschke, C. (2012). Point of gaze estimation through corneal surface reflection in an active illumination environment. In *Proceedings of the 12th European Conference on Computer Vision (ECCV)* (pp. 159–172). doi:10.1007/978-3-642-33709-3_12

Nakazawa, A., Nitschke, C., & Nishida, T. (2015). Non-calibrated and realtime human view estimation using a mobile corneal imaging camera. In *Proceedings of the International Workshop on Wearable and Ego-vision Systems for Augmented Experience (WEsAX)* (pp. 1–6). doi:10.1109/ICMEW.2015.7169846

Nakazawa, A., Nitschke, C., & Nishida, T. (2016). Registration of eye reflection and scene images using an aspherical eye model. *Journal of the Optical Society of America A, 33*(11), 2264–2276.

Nayar, S. K., & Peri, V. (1999). Folded catadioptric cameras. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 217–223). doi:10.1109/CVPR.1999.784632

Nene, S. A., & Nayar, S. K. (1998). Stereo with mirrors. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 1087–1094). doi:10.1109/ICCV.1998.710852

Newcombe, R. A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A. J.,… Fitzgibbon, A. (2011). Kinectfusion: Real-time dense surface mapping and tracking. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 127–136). doi:10.1109/ISMAR.2011.6092378

Nishino, K., Belhumeur, P. N., & Nayar, S. K. (2005). Using eye reflections for face recognition under varying illumination. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 519–526). doi:10.1109/ICCV.2005.243

Nishino, K., & Nayar, S. K. (2004). Eyes for relighting. *ACM Transactions on Graphics, 33*(6), 704–711.

Nishino, K., & Nayar, S. K. (2006). Corneal imaging system: Environment from eyes. *International Journal of Computer Vision, 70*(1), 23–40.

Nitschke, C. (2011). *Image-based eye pose and reflection analysis for advanced interaction techniques and scene understanding* (Unpublished PhD thesis). Graduate School of Information Science and Technology, Osaka University, Osaka, Japan. Retrieved from http://hdl.handle.net/11094/2797

Nitschke, C., & Nakazawa, A. (2012). Super-resolution from corneal images. In *Proceedings of the British Machine Vision Conference (BMVC)* (pp. 1–12). doi:10.5244/C.26.22

Nitschke, C., Nakazawa, A., & Nishida, T. (2013a). I see what you see: Point of gaze estimation from corneal images. In *Proceedings of the IAPR Asian Conference on Pattern Recognition (ACPR)* (pp. 298–304). doi:10.1109/ACPR.2013.84

Nitschke, C., Nakazawa, A., & Takemura, H. (2009a). Display-camera calibration from eye reflections. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 1226–1233). doi:10.1109/ICCV.2009.5459330

Nitschke, C., Nakazawa, A., & Takemura, H. (2009b). Eye reflection analysis and application to display-camera calibration. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)* (pp. 3449–3452). doi:10.1109/ICIP.2009.5413852

Nitschke, C., Nakazawa, A., & Takemura, H. (2011a). Display-camera calibration using eye reflections and geometry constraints. *Computer Vision and Image Understanding, 115*(6), 835–853.

Nitschke, C., Nakazawa, A., & Takemura, H. (2011b). Practical display-camera calibration from eye reflections using coded illumination. In *Proceedings of the IEEE Asian Conference on Pattern Recognition (ACPR)* (pp. 550–554). doi:10.1109/ACPR.2011.6166661

Nitschke, C., Nakazawa, A., & Takemura, H. (2013b). Corneal imaging revisited: An overview of corneal reflection analysis and applications. *IPSJ Transactions on Computer Vision and Applications, 5*, 1–18.

Ohno, T. (2006). One-point calibration gaze tracking method. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA)* (p. 34). doi:10.1145/1117309.1117318

Ohno, T., Mukawa, N., & Yoshikawa, A. (2002). Freegaze: A gaze tracking system for everyday gaze interaction. In *Proceedings of the ACM Symposium on Eye Tracking Research and Applications (ETRA)* (pp. 125–132). doi:10.1145/507072.507098

Pamplona, V. F., Mohan, A., Oliveira, M. M., & Raskar, R. (2010). NETRA: Interactive display for estimating refractive errors and focal range. *ACM Transactions on Graphics, 29*(4), 77:1–77:8.

Pamplona, V. F., Oliveira, M. M., & Baranoski, G. V. G. (2009). Photorealistic models for pupil light reflex and iridal pattern deformation. *ACM Transactions on Graphics, 28*(4), 1–12.

Pamplona, V. F., Passos, E. B., Zizka, J., Oliveira, M. M., Lawson, E., Clua, E., & Raskar, R. (2011). CATRA: Interactive measuring and modeling of cataracts. *ACM Transactions on Graphics, 30*(4), 47:1–47:8.

Park, H. M., Lee, S. H., & Choi, J. S. (2008). Wearable augmented reality system using gaze interaction. In *Proceedings of the IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 175–176). doi:10.1109/ISMAR.2008.4637353

Pharr, M., & Humphreys, G. (2004). *Physically based rendering: From theory to implementation*. San Francisco, CA: Morgan Kaufmann Publishers Inc..

Pires, B. R., Devyver, M., Tsukada, A., & Kanade, T. (2013). Unwrapping the eye for visible-spectrum gaze tracking on wearable devices. In *IEEE Workshop on Applications of Computer Vision (WACV)* (pp. 369–376). doi:10.1109/WACV.2013.6475042

Pires, B. R., Hwangbo, M., Devyver, M., & Kanade, T. (2013). Visible spectrum gaze tracking for sports. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 1005–1010). doi:10.1109/CVPRW.2013.146

Plopski, A. (2016). *Improving optical-see-through experience through corneal imaging* (Unpublished PhD thesis). Graduate School of Information Science and Technology, Osaka University, Osaka, Japan. Retrieved from http://hdl.handle.net/11094/55861

Plopski, A., Itoh, Y., Nitschke, C., Kiyokawa, K., Klinker, G., & Takemura, H. (2015a). Corneal-imaging calibration for optical see-through head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics, 21*(4), 481–490.

Plopski, A., Kiyokawa, K., Takemura, H., & Nitschke, C. (2014). Corneal imaging in localization and HMD interaction. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 397–400). doi:10.1109/ISMAR.2014.6948505

Plopski, A., Nitschke, C., Kiyokawa, K., Schmalstieg, D., & Takemura, H. (2015b). Hybrid eye tracking: Combining iris contour and corneal imaging. In *Proceedings of the International Conference on Artificial Reality and Telexistence (ICAT)* (pp. 1–8). doi:10.2312/egve.20151327

Plopski, A., Orloski, J., Itoh, Y., Nitschke, C., Kiyokawa, K., & Klinker, G. (2016). Automated spatial calibration of HMD systems with unconstrained eye-cameras. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR)* (pp. 1–6). doi:10.1109/ISMAR.2016.16

Porta, M. (2002). Vision-based user interfaces: Methods and applications. *International Journal of Human-Computer Studies, 57*(1), 27–73.

Priamikov, A. & Triesch, J. (2014). Openeyesim—a platform for biomechanical modeling of oculomotor control. In *Proceedings of the Joint IEEE International Conferences on Development and Learning and Epigenetic Robotics (ICDL-Epirob)* (pp. 394–395). doi:10.1109/DEVLRN.2014.6983013

Proença, H. (2013). Iris recognition in the visible wavelength. In M. J. Burge, & K. W. Bowyer (Eds.), *Handbook of iris recognition, advances in computer vision and pattern recognition* (pp. 151–169). London: Springer.

Reale, M., Canavan, S., Yin, L., Hu, K., & Hung, T. (2011). A multi-gesture interaction system using a 3-D iris disk model for gaze estimation and an active appearance model for 3-D hand pointing. *IEEE Transactions on Multimedia, 13*(3), 474–486.

Reale, M., Hung, T., & Yin, L. (2010). Viewing direction estimation based on 3D eyeball construction for HRI. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 24–31). doi:10.1109/CVPRW.2010.5543784

Remington, L. A. (2004). *Clinical anatomy of the visual system* (2nd ed.). St. Louis, MO: Butterworth-Heinemann.

Rolland, J. P., Holloway, R. L., & Fuchs, H. (1995). Comparison of optical and video see-through, head-mounted displays. In *Proceedings of the Photonics for Industrial Applications (SPIE)* (pp. 293–307). doi:10.1117/12.197322

Ruhland, K., Andrist, S., Badler, J., Peters, C., Badler, N., Gleicher, M.,… Mcdonnell, R. (2014). Look me in the eyes: A survey of eye and gaze animation for virtual agents and artificial systems. In *Proceedings of the Eurographics State-of-the-Art Report* (pp. 69–91). doi:10.2312/egst.20141036

Ryan, W. J., Woodard, D. L., Duchowski, A. T., & Birchfield, S. T. (2008). Adapting starburst for elliptical iris segmentation. In *Proceedings of the IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)* (pp. 1–7). doi:10.1109/BTAS.2008.4699340

Safaee-Rad, R., Tchoukanov, I., Smith, K. C., & Benhabib, B. (1992). Three-dimensional location estimation of circular features for machine vision. *IEEE Transactions on Robotics and Automation, 8*(5), 624–640.

Sagar, M. A., Bullivant, D., Mallinson, G. D., & Hunter, P. J. (1994). A virtual environment and model of the eye for surgical simulation. In *Proceedings of the ACM SIGGRAPH '94* (pp. 205–212). doi: 10.1145/192161.192200

Salvi, J., Fernandez, S., Pribanic, T., & Llado, X. (2010). A state of the art in structured light patterns for surface profilometry. *Pattern Recognition, 43*(8), 2666–2680.

Schindler, G. (2008). Photometric stereo via computer screen lighting for realtime surface reconstruction. In *Proceedings of the International Symposium on 3D Data Processing, Visualization and Transmission (3DPVT)* (pp. 1–6). Retrieved from https://www.cc.gatech.edu/~dellaert/ftp/3DPVT08.pdf

Schnieders, D., Fu, X., & Wong, K.-Y. K. (2010). Reconstruction of display and eyes from a single image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1442–1449). doi:10.1109/CVPR.2010.5539799

Schnieders, D., Wong, K.-Y. K., & Dai, Z. (2009). Polygonal light source estimation. In *Proceedings of the Asian Conference on Computer Vision (ACCV)* (pp. 96–107). doi:10.1007/978-3-642-12297-2_10

Semple, J. G. & Kneebone, G. T. (1952). *Algebraic projective geometry*. Oxford: Oxford University Press.

Shih, S.-W., Wu, Y.-T., & Liu, J. (2000). A calibration-free gaze tracking technique. In *Proceedings of the IEEE International Conference on Pattern Recognition (ICPR)* (pp. 4201–4204). doi:10.1109/ICPR.2000.902895

Smith, J. D., Vertegaal, R., & Sohn, C. (2005). Viewpointer: Lightweight calibration-free eye tracking for ubiquitous handsfree deixis. In *Proceedings of the ACM Symposium on User Interface Software and Technology (UIST)* (pp. 53–61). doi:10.1145/1095034.1095043

Snell, R. S., & Lemp, M. A. (1997). *Clinical anatomy of the eye* (2nd ed.). Malden, MA: Blackwell Publishing.

Somppi, S., Törnqvist, H., Hänninen, L., Krause, C., & Vainio, O. (2012). Dogs do look at images: Eye tracking in canine cognition research. *Animal Cognition, 15*(2), 163–174.

Stampe, D. M. (1993). Heuristic filtering and reliable calibration methods for video-based pupil-tracking systems. *Behavior Research Methods, Instruments, and Computers, 25*(2), 137–142.

Strelow, D., Mishler, J., Koes, D., & Singh, S. (2001). Precise omnidirectional camera calibration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 689–694). doi:10.1109/CVPR.2001.990542

Sturm, P. (2002). Mixing catadioptric and perspective cameras. In *Proceedings of the IEEE Workshop on Omnidirectional Vision (OMNIVIS)* (pp. 37–44). doi:10.1109/OMNVIS.2002.1044489

Sturm, P., & Bonfort, T. (2006). How to compute the pose of an object without a direct view? In *Proceedings of the Asian Conference on Computer Vision (ACCV)* (pp. 21–31). doi:10.1007/11612704_3

Sturm, P., Ramalingam, S., Tardif, J.-P., Gasparini, S., & Barreto, J. (2011). Camera models and fundamental concepts used in geometric computer vision. *Foundations and Trends in Computer Graphics and Vision, 6*(1–2), 1–183.

Sugano, Y., Matsushita, Y., & Sato, Y. (2013). Appearance-based gaze estimation using visual saliency. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 35*(2), 329–341.

Sugano, Y., Matsushita, Y., Sato, Y., & Koike, H. (2008). An incremental learning method for unconstrained gaze estimation. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 656–667). doi:10.1007/978-3-540-88690-7_49

Svoboda, T. & Pajdla, T. (2002). Epipolar geometry for central catadioptric cameras. *International Journal of Computer Vision, 49*(1), 23–37.

Swaminathan, R., Grossberg, M. D., & Nayar, S. K. (2006). Non-single viewpoint catadioptric cameras: Geometry and analysis. *International Journal of Computer Vision, 66*(3), 211–229.

Swartz, T., Marten, L., & Wang, M. (2007). Measuring the cornea: the latest developments in corneal topography. *Current Opinion in Ophthalmology, 18*(4), 325–333.

Świrski, L. & Dodgson, N. (2014). Rendering synthetic ground truth images for eye tracker evaluation. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA)* (pp. 219–222). doi:10.1145/2578153.2578188

Takemura, K., Kimura, S., & Suda, S. (2014a). Estimating point-of-regard using corneal surface image. In *Proceedings of the ACM Symposium on Eye Tracking Research and Applications (ETRA)* (pp. 251–254). doi:10.1145/2578153.2578197

Takemura, K., Yamakawa, T., Takamatsu, J., & Ogasawara, T. (2014b). Estimation of a focused object using a corneal surface image for eye-based interaction. *Journal of Eye Movement Research, 7*(3), 1–9.

Tan, T., He, Z., & Sun, Z. (2010). Efficient and robust segmentation of noisy iris images for non-cooperative iris recognition. *Image and Vision Computing, 28*(2), 223–230.

Tarini, M., Lensch, H. P. A., Goesele, M., & Seidel, H.-P. (2005). 3D acquisition of mirroring objects using striped patterns. *Graphical Models, 67*(4), 233–259.

Tian, J., & Ma, K.-K. (2011). A survey on super-resolution imaging. *Signal, Image and Video Processing, 5*(3), 329–342.

Tomasi, C., & Manduchi, R. (1998). Bilateral filtering for gray and color images. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 839–846). doi:10.1109/ICCV.1998.710815

Tsukada, A., & Kanade, T. (2012). Automatic acquisition of a 3D eye model for a wearable first-person vision device. In *Proceedings of the Symposium on Eye Tracking Research and Applications (ETRA)* (pp. 213–216). doi:10.1145/2168556.2168597

Tsukada, A., Shino, M., Devyver, M., & Kanade, T. (2011). Illumination-free gaze estimation method for first-person vision wearable device. In *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCVW)* (pp. 2084–2091). doi:10.1109/ICCVW.2011.6130505

Tsumura, N., Dang, M. N., Makino, T., & Miyake, Y. (2003). Estimating the directions to light sources using images of eye for reconstructing 3D human face. In *Proceedings of the IS&T/SID Color Imaging Conference (CIC)* (pp. 77–81). Retrieved from http://www.ingentaconnect.com/content/ist/cic/2003/00002003/00000001/art00014

Tuceryan, M., Genc, Y., & Navab, N. (2002). Single-point active alignment method (SPAAM) for optical see-through HMD calibration for augmented reality. *Presence, 11*(3), 259–276.

Turk, M. (2004). Computer vision in the interface. *Communications of the ACM, 47*(1), 60–67.

Tuytelaars, T., & Mikolajczyk, K. (2008). Local invariant feature detectors: A survey. *Foundations and Trends in Computer Graphics and Vision, 3*(3), 177–280.

Tweed, D., & Vilis, T. (1990). Geometric relations of eye position and velocity vectors during saccades. *Vision Research, 30*(1), 111–127.

van de Kraats, J., & van Norren, D. (2008). Directional and nondirectional spectral reflection from the human fovea. *Journal of Biomedical Optics, 13*(2), 024010–024010–13.

Vanderportaele, B. (2006). *Contributions à la vision omnidirectionnelle: etude, conception et etalonnage de capteurs pour l'acquisition d'images et la modélisation 3D.* (Unpublished PhD thesis), Institut National Polytechnique de Toulouse, Toulouse. Retrieved from http://ethesis.inp-toulouse.fr/archive/00000372/

Vezhnevets, V., & Degtiareva, A. (2003). Robust and accurate eye contour extraction. In *Proceedings of the Graphicon* (pp. 81–84). Moscow: GraphiCon Scientific Society. Retrieved from http://graphicon.ru/html/2003/Proceedings/Technical/paper492.pdf

Villanueva, A., & Cabeza, R. (2007). Models for gaze tracking systems. *EURASIP International Journal of Image and Video Processing, 2007*(3), 1–16.

Villanueva, A., & Cabeza, R. (2008). A novel gaze estimation system with one calibration point. *IEEE Transactions on Systems, Man, and Cybernetics Part B: Cybernetics, 38*(4), 1123–1138.

Villanueva, A., Daunys, G., Hansen, D. W., Böhme, M., Cabeza, R., Meyer, A., & Barth, E. (2009). A geometric approach to remote eye tracking. *Universal Access in the Information Society, 8*(4), 241–257.

Wang, H., Lin, S., Liu, X., & Kang, S. B. (2005). Separating reflections in human iris images for illumination estimation. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)* (pp. 1691–1698). doi:10.1109/ICCV.2005.215

Wang, H., Lin, S., Ye, X., & Gu, W. (2008). Separating corneal reflections for illumination estimation. *Neurocomputing, 71*(10–11), 1788–1797.

Wang, J.-G. & Sung, E. (2001). Gaze determination via images of irises. *Image and Vision Computing, 19*(12), 891–911.

Wang, J.-G. & Sung, E. (2002). Study on eye gaze estimation. *IEEE Transactions on Systems, Man, and Cybernetics Part B: Cybernetics, 32*(3), 332–350.

Watson, A. B. (1986). Temporal sensitivity (time dependent human perception of visual stimuli. In K. Boff, L. Kaufman & J. Tomas (Eds.), *Handbook of perception and human performance* (pp. 6-1–6-43). New York, NY: Wiley Interscience.

Wood, E., Baltrusaitis, T., Zhang, X., Sugano, Y., Robinson, P., & Bulling, A. (2015). *Rendering of eyes for eye-shape registration and gaze estimation.* Retrieved from http://arxiv.org/abs/1505.05916v1

Wu, H., Chen, Q., & Wada, T. (2005). Visual direction estimation from a monocular image. *IEICE Transactions on Information and Systems, E88-D*(10), 2277–2285.

Wu, H., Kitagawa, Y., Wada, T., Kato, T., & Chen, Q. (2007). Tracking iris contour with a 3D eye-model for gaze estimation. In *Proceedings of the Asian Conference on Computer Vision (ACCV)* (pp. 688–697). doi:10.1007/978-3-540-76386-4_65

Würz-Wessel, A. (2003). *Free-formed Surface Mirrors in Computer Vision Systems* (Unpublished PhD thesis). Universität Tübingen, Tübingen. Retrieved from http://hdl.handle.net/10900/48482

Yagi, Y. (1999). Omnidirectional sensing and its applications. *IEICE Transactions on Information and Systems, E82-D*(3), 568–579.

Yamazoe, H., Utsumi, A., Yonezawa, T., & Abe, S. (2008). Remote and head-motion-free gaze tracking for real environments with automated head-eye model calibrations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 1–6). doi:10.1109/CVPRW.2008.4563184

Ying, J., Wang, B., & Shi, M. (2012). Anterior corneal asphericity calculated by the tangential radius of curvature. *Journal of Biomedical Optics, 17*(7), 075005–1–075005–9.

Yoo, D. H., & Chung, M. J. (2005). A novel non-intrusive eye gaze estimation using cross-ratio under large head motion. *Computer Vision and Image Understanding, 98*(1), 25–51.

Young, L., & Sheena, D. (1975). Survey of eye movement recording methods. *Behavior Research Methods and Instrumentation, 7*(5), 397–429.

Zheng, Y., Ma, W., & Liu, Y. (2008). Another way of looking at monocular circle pose estimation. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)* (pp. 861–864). doi:10.1109/ICIP.2008.4711891

The Wiley Handbook of Human Computer Interaction

The Wiley Handbook of Human Computer Interaction

Volume 2

Edited by

Kent L. Norman and Jurek Kirakowski

WILEY Blackwell

This edition first published 2017

Registered Offices
John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Office
The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Library of Congress Cataloging-in-Publication Data Is Available

9781118976135 – Hardback
9781118977262 – e-PDF
9781118977279 – e-Pub

Cover Design: Wiley
Cover Image: © Andrea Danti/Shutterstock

Set in 10/12pt Galliard by SPi Global, Pondicherry, India.

10 9 8 7 6 5 4 3 2 1

Contents

Notes on Contributors

June Ahn is an associate professor of learning sciences and educational technology at New York University, Steinhardt School of Culture, Education, and Human Development. He conducts research on the design, implementation, and evaluation of learning technologies. He is interested in designing and understanding sociotechnical systems—or how social, cultural, and institutional factors intersect with new technologies to create enhanced and equitable learning opportunities for all learners. Dr. Ahn has conducted research on the use of social media for learning, alternative reality games and playful learning, and the use of analytics for research-practice partnerships with school districts. This work has been supported by grants from the National Science Foundation and the Institute of Education Sciences.

Claudia Savina Bianchini obtained a PhD in linguistics in 2012, and co-tutored in France (University of Paris VIII) and Italy (University Studi Perugia). She started teaching linguistics, specifically sign language linguistics, in 2012 at the University of Poitiers (France), and became a tenured associate professor in 2013. The subjects of her works—mainly focused on systems for transcription and writing of sign languages—range from the "systematic analysis" of SignWriting to the "creation" of Typannot (a brand-new writing system for sign languages), passing by the typographic representation of glyphs and the phonological analysis of sign language. She has authored 22 scientific papers, many in international refereed journals; her papers have been cited over 100 times, and her works have been awarded four prizes. She has presented numerous contributions at national and international conferences, including as an invited speaker, and has peer reviewed many papers. She has participated, and still participates, in various research programs at national level. She belongs to the FoReLL (Formes et Représentations en Linguistique et Littérateur) laboratory of the University of Poitiers. She has traveled extensively in Europe and the Americas. She speaks four vocal languages and uses two sign languages (LIS and LSF).

Gary Borchardt is a research affiliate at the MIT Computer Science and Artificial Intelligence Laboratory and holds a PhD (1992) in computer science from MIT. His research focuses on commonsense reasoning about events, question answering, and language-based representation of knowledge.

Fabrizio Borgia has a PhD in computer science, awarded by Université Toulouse III—Paul Sabatier, in international partnership with the Sapienza University of Rome. His main scientific interests are in the field of HCI and image processing. Accessibility for people with special needs and user interface evaluation techniques are among its main drivers in the HCI field. His image-processing research is mainly focused on optical character recognition techniques. Fabrizio also specialized in image processing for satellite imagery, and currently works as an Earth Observation Data Dissemination Engineer at the European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT).

Anthony Brooks is associate professor at Aalborg University, Denmark, where he is director/founder of the "SensoramaLab." He was a founding team-member of Medialogy education, where he is section leader, lecturer, coordinator, supervisor, and study board member. Originating from Wales, he created a bespoke volumetric invisible sensing system for unencumbered gesture-control of media to stimulate interactions and human performance potential. In the mid-1990s, he renamed his maturing body of work as "SoundScapes." Virtual worlds for creative expression, games for serious applications, and virtual interactive space (VIS) are core aspects of his human-centered research, targeting alternative channeling of biofeedback to realize human afferent-efferent neural feedback loop closure as a supplement for traditional rehabilitation, healthcare, and quality-of-life interventions. He originated the ZOOM (Zone of Optimized Motivation) model for in-action intervention and an on-action analysis and assessment model under a custom-made synthesized methodology of action research integrated with hermeneutics—resulting in the recursive reflection model. He has approximately 200 publications, including five books published by Springer, among which are *Recent Advances in Technologies for Inclusive Well-Being: From Worn to Off-body Sensing, Virtual Worlds, and Games for Serious Applications* (2017) and *Interactivity, Game Creation, Design, Learning, and Innovation* (2017). His research is responsible for sizable externally funded national and international (European) projects, a game industry startup, and a patented commercial product (Patent US6893407 "Communication Method and Apparatus"). He is acknowledged as a third culture thinker and "world pioneer in digital media and its use with the disabled community." He is an active keynote speaker at international events—the latest being "Smart Digital Futures" in Vilamoura, Algarve, Portugal, and has presented SoundScapes globally (including at the Olympics and Paralympics in 1996 and 2000, and numerous others). He has hosted numerous international conferences as organizing chair. He is Danish representative for UNESCO's International Federation for Information Processing (IFIP) technical committee (WG14) on "Entertainment Computing"—specifically under work groups WG14.7 "Art and Entertainment"; WG14.8 "Serious Games," and WG 14.9 "Game Accessibility." He was appointed by the European Commission as EU expert examiner, rapporteur (Serious Games/Gamification, Human-Computer Confluence, Presence and Future Emerging Technologies) and panel reviewer of funded projects (e.g. Presenccia—an Integrated Project funded under the European Sixth Framework Program, Future and Emerging Technologies (FET), Contract Number 27731.) As vice-chair of the European Alliance for Innovation's Wellbeing (SIB-WellB) Market Trends and Society segment, he is also steering person for the international conference "ArtsIT, Interactivity and Game Creation." He achieved his PhD at the

University of Sunderland, Great Britain. Amongst his many boards/committees, he is active in the Program Board of the International Conference on Universal Access in Human-Computer Interaction (UAHCI) one of the Affiliated Conferences jointly held under one management and one registration in the context of HCI International—The International Conference on Human-Computer Interaction.

Luís Carriço is an associate professor at the Informatics Department, Faculty of Sciences, University of Lisbon. He is the director of LaSIGE, a CS research unit, and senior member of the Human Computer Interaction and Multimedia Group. He holds a PhD (2000) in electronic and computers engineering from IST, Technical University of Lisbon. He has published over 200 refereed technical/scientific papers, addressing mobile interaction design, support for cognitive impairments, and mobile and Web accessibility.

Tiziana Catarci is full professor in computer science and engineering at the Sapienza University of Roma and director of the ECONA Interuniversity Research Centre on the Cognitive Processing in Natural and Artificial Systems and of the Sapienza Design Research Center. During 2010–2014 she was Sapienza vice-rector for technologies and infrastructures and during 2009–2014 she was president of the Sapienza ICT service center. Tiziana Catarci's main research interests are in theoretical and application-oriented aspects of visual information access, human-computer interaction, user interfaces, usability, smart environments, information visualization, data management, data quality, and digital libraries. On these topics she has published over 200 papers in international journals, conferences, and workshop proceedings, and 20 books. Her contribution can be regarded as one of the first and most significant examples of deep analysis and formalization of the interaction between the user and the database, which takes into consideration both usability issues and language related aspects. Dr. Catarci is the editor-in-chief of the *ACM Journal of Data and Information Quality*. She is regularly on the programming committees of the main database and HCI conferences and is associate editor of the *World Wide Web Journal* and the *Journal of Data Semantics*. She has been cochair of several important international conferences in both database and human-computer interaction areas. In particular, in 2008, she has been the cochair of the 2008 edition of the largest and most important conference on HCI, ACM CHI. In 2003, she has been the recipient of the IBM Eclipse Innovation Award. In 2016 she has been nominated member of the prestigious European Academy of Sciences and Arts. In 2016 she has been included among the "100 Women for Science" project - http://www.100esperte.it/. In 2017, she received the Levi Montalcini Association award for the "diffusion of scientific culture among younger generations."

Ben Challis is a principal lecturer in contemporary arts at Manchester Metropolitan University where he lectures in composition and music technology. With research interests that embrace the notion of design for all within music performance, he has worked on various projects that explore alternative modes of interaction with sound and music for people with specific individual needs. As a performer, he works with these same technologies, exploring their creative and expressive potential within free improvisation. As a composer/sound designer he has created works for dance, theater, short films and computer games.

Eugene Cho is a PhD student in mass communications at the Pennsylvania State University. She holds a master's degree in communication from Seoul National University. Her research pertains to social psychological effects of digital media usage, with a particular interest in how specific technological affordances of digital media alter individuals' perceptions, attitudes, and behaviors.

Jun-Dong Cho received a BS degree in electronic engineering from Sungkyunkwan University, Seoul, Korea in 1980; an MS degree from Polytechnic University, Brooklyn, New York in 1989, and a PhD degree from Northwestern University, Evanston, IL in 1993, both in computer science. He was a senior CAD engineer at Samsung Electronics. He is now in several faculty positions in Sungkyunkwan University: professor in the Department of Electronic Engineering (director of VADA Lab. http://vada.skku.ac.kr), chair of the Department of Human ICT Convergence (Director of H-Lab) http://humanict.skku.edu, and adjunct professor, Department of Interaction Science. He received the Best Paper award at the 1993 Design Automation Conference. He has been an IEEE senior member since April 1996. He received the Research Achievement Award and Best Convergence Research Award from the Ministry of Knowledge and Economics in October, 2012 and Ministry of Science, ICT and Future Planning in July 2014, respectively. His recent research interest is enhancing the usability of smart wearable healthcare devices.

Young-Joon Cho received a BS in microbiology and a PhD in biological science from Seoul National University. He is currently a research scientist in the Korea Polar Research institute. His research interests include genome and transcriptome analysis.

Tamara Clegg is an assistant professor in the College of Information Studies and the Department of Teaching and Learning, Policy and Leadership at the University of Maryland. Prior to her current position, she received her PhD in computer science from the Georgia Institute of Technology and was a Computing Innovations Fellow in the College of Information Studies at the University of Maryland. Her work focuses on developing technology to support life-relevant learning, where children engage in science in the context of achieving personally relevant goals. She seeks to understand ways such learning environments and technologies support scientific disposition development. Tamara's work is funded by the National Science Foundation, the Institute of Museum and Library Studies, and Google.

David Costa is a PhD student at the Faculty of Sciences of the University of Lisbon and an HCI researcher at LaSIGE. David Costa currently holds a master's degree, obtained in 2011, for his research on multimodal output adaptation. Currently researching in the field of mobile accessibility for the blind, his goal is to design a body interaction model to describe and then ease interaction with mobile devices for this target group.

Anna L. Cox is a professor of HCI and deputy director of UCLIC. Her research focuses on interactions between people and digital technologies in both work contexts (e.g. crowdsourced data entry) and online leisure activities (e.g. games, crowdsourced citizen science). She currently leads the GetAMoveOn Network+and has been a principal/coinvestigator for several other research projects, including Digital

Epiphanies, Citizen Cyberlab, and Open 3D. She has published over 140 refereed journal and conference publications. Notably, she coedited the first text book on *Research Methods for Human-Computer Interaction* (Cairns & Cox, 2008). Together, Charlene Jennett and Anna have co-authored several publications on the topic of HCI and digital citizen science, including: *Motivations, Learning and Creativity in Online Citizen Science* (2016); *Designing for Dabblers and Deterring Drop-outs in Citizen Science* (2014); *I want to be a Captain! I want to be a Captain! Gamification in the Old Weather Citizen Science Project* (2013), and *Do Games Attract or Sustain Engagement in Citizen Science? A Study of Volunteers' Motivations* (2013). They also coauthored the journal paper "Measuring and defining the experience of immersion in games" (2008), which has received over 700 citations to date.

Maria De Marsico is associate professor in the Department of Computer Science at Sapienza University, Rome. Her scientific interests focus on image processing and human computer interaction. Regarding the first one, she is especially interested in multimodal interaction, accessibility for users with special needs, and advanced techniques for personalized distance learning. Regarding the second one, she works on biometric recognition, including face, iris, gate, and multimodal recognition. She is associate editor of *Pattern Recognition Letters*, and area editor of the IEEE Biometrics Compendium. She published about 150 scientific works in international journals, conferences, and book chapters. She has been a member of many technical program committees and is referee for several top journals. She is program chair for the International Conference on Pattern Recognition Applications and Methods since 2013, and has been Program Chair of CHItaly 2015, the biannual conference of the Italian chapter of ACM SIGCHI.

Carlos Duarte is an assistant professor at the Informatics Department, Faculty of Sciences, University of Lisbon. He is also a senior researcher in the Human Computer Interaction and Multimedia Group of LaSIGE. He holds a PhD (2008) in informatics from the University of Lisbon. He has published over 100 refereed papers. His research interests focus on promoting the access to interactive systems for populations with special needs through combinations of adaptation and multimodality.

Alistair Edwards is a senior lecturer in computer science and member of the HCI research group at the University of York. His main research interests are around multimodal interaction. Much of this work is motivated by the needs of exceptional users who can have difficulties in accessing conventional interfaces. This goes back to his PhD work at the Open University on making graphical interfaces accessible to blind users. His research interests are generally clustered around this topic and include access to mathematics, computers in theater, speech-based interaction, sonification, tactile interaction, and use of computers by older people.

Sue Felshin is a research scientist in the InfoLab Group at MIT's Computer Science and Artificial Intelligence Laboratory and has a degree in cognitive science from the Massachusetts Institute of Technology. Her research interests include natural language understanding and generation, particularly with respect to semantics, discourse and dialog, and HCI.

Jennifer Golbeck is an associate professor in the College of Information Studies at the University of Maryland. She is director of the Social Intelligence Lab where she studies the intersection of social media, HCI, and artificial intelligence. Golbeck is the author of *Analyzing the Social Web* (2013), and *Introduction to Social Media Investigation* (2015). Her PhD is in computer science from the University of Maryland, with bachelors and masters degrees from the University of Chicago.

Paul A. Green is a research professor in the Driver Interface Group at the University of Michigan Transportation Research Institute (UMTRI) and an adjunct professor/research professor of industrial and operations engineering. He has a PhD in industrial and operations engineering and psychology from the University of Michigan and is a BCPE board-certified human factors professional. He teaches courses on automotive human factors, human-computer interaction, and is the leader of the human factors engineering short course, the flagship continuing education course in the field, now in its 58th year. His research concerns driver interfaces, driver workload and distraction, and other topics. Much of his current work relates to predicting driver performance, research that makes extensive use of instrumented cars and a driving simulator. He is the author or co-author of more than 300 publications (www.umich.edu/~driving). He is the lead author of four SAE recommended practices, including SAE J2365, which estimates task times for using in-vehicle interfaces, and SAE J2944, which defines driving performance measures and statistics.

Eve Griffin holds a PhD in social psychology from University College Cork. Her research is concerned with the processes of social exchange in Internet-mediated communication, with a particular interest in impression formation in nonverbal communication. A significant aspect of this work was the development of a standardized scale to measure attitudes to online relationships, which can be applied across different fields of usability, and psychology.

Jeong-Ki Hong received his BS degree in mechanical engineering from Yonsei University in Seoul, Korea, in 2012, and his MS degree in convergence science from the Sungkyunkwan University, Seoul, Korea, in 2015. He mainly studied on HCI including topics such as a haptic or ambient feedback, gesture sensing, quantified-self projects and gamification systems. He now is a research engineer at Hyundai Motors.

Tharon Howard teaches in the Master of Arts in professional communication program and the rhetorics, communication, and information design doctoral program at Clemson University. In 2017, he received the Society of Technical Communication's Rainey Award for Excellence in Research from the Society of Technical Communication and in 2011 the J. R. Gould Award for Excellence in Teaching Technical Communication. As director of the Clemson University Usability Testing Facility, Howard has conducted sponsored research aimed at improving and creating new software interfaces, online document designs, and information architecture for clients including Pearson Higher Education, IBM, NCR Corp, and AT&T. For his work promoting the importance of usability in both industry and technical communication, Howard was awarded the Usability Professionals Association's very first "Extraordinary Service Award." Howard is the Series Editor for the Routledge-ATTW Series on Technical and Professional Communication. He also serves as the production manager for Clemson's Center for Electronic and Digital Publishing where—in addition to producing scholarly journals, books, fliers, and brochures—he teaches graduate students to create and maintain digital publications and websites. He designed and directed Clemson's Multimedia Authoring Teaching and Research Facility, where faculty and graduate students in architecture, arts, and humanities learn to develop fully interactive, stand-alone multimodal productions and experiment with emerging instructional technologies, augmented reality devices, and interface designs. Howard has written or coauthored four books, the most recent being *Design to Thrive: Creating Online Communities and Social Networks That Last* published in 2010 by Morgan Kauffman.

Hung-Hsuan Huang is an associate professor at College of Information Science and Engineering, Ritsumeikan University. He received his BS and M.E. from National Cheng-Chi University and National Taiwan University in 1998 and 2000, respectively. In 2009, he received his PhD from the Graduate School of Informatics, Kyoto University, Japan. He then worked as a postdoctoral researcher at the Department of Computer and Information Science, Seikei University, until he transferred to Ritsumeikan University as an assistant professor in 2010. His research interests are in applied artificial intelligence and human-computer interaction fields. Currently he intensively focused his works on embodied conversational agents (ECAs) and multi-modal interaction. He is a member of ACM, JSAI, IPSJ, HIS, IEICE, VRSJ, and TAAI.

Charlene Jennett is a Research Associate at the University College London Interaction Centre (UCLIC). She received her doctorate in HCI from UCLIC in 2010. Her research interests include HCI, digital games, and citizen science. During 2012–2016, she worked on the *Citizen Cyberlab* project (EU funded), where she investigated volunteers' motivations, learning, and creativity in digital citizen science. Currently she works on the *Open 3D* project (Engineering and Physical Sciences Research Council funded), where she continues to investigate users' experiences of digital citizen science.

Mary Joyce completed her PhD in University College Cork, Ireland. Her research focuses on the measurement of Internet attitudes and self-efficacy, and the use of such scales in usability evaluation. Mary is particularly interested in using these scales to examine individual differences in usability studies. She recently published her work on the development of the General Internet Attitude Scale and is pursuing further research on the importance of self-efficacy measurement in usability evaluation.

Boris Katz is a principal research scientist and head of the InfoLab group at MIT's Computer Science and Artificial Intelligence Laboratory. He also serves as a leader of the Vision and Language Thrust of the MIT-Harvard Center for Brains, Minds, and Machines. His research interests include natural language understanding and generation, HCI, and integration of language, vision, and robotics. Boris Katz is the creator of the START information access system, the first question-answering system on the Web. He is a founding member of the Open Advancement of Question Answering consortium, where he contributed several technical ideas incorporated into IBM's Watson system, which in 2011 defeated the all-time human champion on the quiz show *Jeopardy!* Technology created in Katz's InfoLab was a major inspiration for the development of Apple's personal assistant, Siri.

Bohyoung Kim received the BS and MS degrees in computer science and the PhD degree in computer science and engineering from Seoul National University, Seoul, Korea in 1995, 1997, and 2001, respectively. She is currently an assistant professor in the Division of Biomedical Engineering, Hankuk University of Foreign Studies. Her research interests include computer graphics, volume visualization, medical imaging, and information visualization.

Stephen Kimani is a senior faculty staff and the director of the School of Computing and Information Technology at Jomo Kenyatta University (JKUAT) in Kenya. He was previously a researcher at CSIRO (Australia), and a postdoctoral researcher at the Sapienza University of Rome. He holds a PhD in computer engineering (University of Rome "La Sapienza," 2004) and an MSc in advanced computing (University of Bristol, 1998). His main research interest is in human-computer interaction (HCI). In particular, as HCI relates to areas / aspects such as user interfaces, usability, accessibility, visualization, visual information access, visual data mining, digital libraries, and ubiquitous computing. He has published widely in book chapters, journal articles, and international conferences and workshops in computer science. He has served on various program committees of and reviewing for international journals, conferences, and workshops. Furthermore, he has been involved in organizing sessions at international conferences and workshops. He has also been involved in proposing and running projects in computer science, including Italian and European research projects. He is actively involved in leading, developing, and coordinating research teams; teaching classes at undergraduate and postgraduate levels; and overseeing / supervising postgraduate student projects.

Bongkyung Ko is a software engineer in Seoul. He graduated from Seoul National University with an MS in HCI. During his studies, he participated in several bioinformatics and information visualization projects. After his graduation, he worked with the development team for STRIKEZON, which is an indoor screen baseball simulator. He has been developing the HANA database management system in SAP since 2016.

Aideen Lawlor is a cyberpsychologist whose research interests include how online communication shapes our identity in the "real world," Internet-related psychopathology, online deception detection and the efficacy of online support groups. Her thesis topic explored Munchausen by Internet with the aim of developing a classifier to aid in its detection. She received her PhD from University College Cork, Ireland.

Jaeyong Lee is a full professor in the Department of Statistics at Seoul National University. He received his PhD in statistics from Purdue University. His research area is Bayesian statistics. In particular, he is interested in nonparametric Bayesian models, high-dimensional statistical models, differential equation models and the frequentist asymptotic properties of Bayesian procedures.

Frank Maurer is the head of the Agile Software Engineering (ASE) group at the University of Calgary. His research interests are immersive analytics, multisurface systems, engineering analytics applications, and agile software methodologies. He served as the principal investigator of the NSERC SurfNet strategic network. The

SurfNet Network was a Canadian research alliance of academic researchers, industry partners, and government collaborators. The goal of SurfNet was to improve the development, performance, and usability of software applications for surface computing environments: nontraditional digital display surfaces including multi-touch screens, tabletops, and wall-sized displays. He served as associate vice-president (research) and special advisor for entrepreneurship and innovation for the University of Calgary. He is cofounder, CTO and board member at VizworX (www.vizworx.com).

John McCarthy is a professor of applied psychology in University College Cork, Ireland, and leads the people and technology research group there. His research is concerned with understanding the influence of emerging social, personal, and work technologies on people's lived experience, and using that understanding to inform design of usable and enriching technologies. With Peter Wright, he has written three books, *Technology as Experience* (MIT Press, 2004), *Experience-Centred Design* (Morgan Claypool, 2010), and *Taking [A]Part* (MIT Press, 2015).

Massimo Mecella, who has a PhD in engineering in computer science, is an associate professor at Sapienza (with a recently earned Italian national full professorship qualification), where he is conducting research in the fields of software architectures, information systems, distributed middleware, service-oriented computing, mobile and pervasive computing, process management, data and process mining, big data analytics, advanced interfaces and HCI, focusing on smart applications, environments and communities. He is author of about 180 papers (h-index = 30). He has been / is currently involved in several European and Italian research projects, and has been the technical manager of the projects WORKPAD and SM4All, coordinated by Sapienza.

Yasser Mohammad is a research engineer at KDDI Research Inc., Japan, and an associate professor at Assiut University, Egypt. His research interests focus on machine understanding and reproduction of human behavior using machine learning and time-series analysis techniques. He received his PhD in 2009 from Kyoto University. He is a co-author of *Conversational Informatics: A Data Intensive Approach* (Springer, 2014) and the lead author of *Data Mining for Social Robotics* (Springer, 2015). He published over 75 international journal and conference papers and coauthored several book chapters on human-robot interaction and data mining. He received the best paper award for IEA/AIE 2008, IEA/AIE 2009, IEEE/SICE SII 2011, ICROS ICCAS 2014, and best presentation award of ICIAE 2015.

Federico Mora is a software engineer and entrepreneur specializing in data science and natural language processing (NLP). He graduated from MIT with a degree in computer science and electrical engineering. While at MIT he carried out research in information retrieval under the guidance of Dr. Boris Katz. His focus is on improvement of HCI with Big Data.

Kellie Morrissey is a Newcastle University Research Fellow the in Open Lab, School of Computing Science, where she leads the digital social care theme. Her research interests lie mainly in the field of HCI with an emphasis on digital social care and digital health. Her most significant work to date was an ethnographic exploration of designing for and with people with dementia living in publicly funded care in the south of

Ireland. She is interested in participatory approaches to research through design with people facing significant challenges, and the potential for digital design to enrich the lives of older populations, particularly those living with, or working with, dementia. She also carries out HCI work in the domain of women's access to healthcare, and the study of online discourse in subversive and marginalized communities.

Xiangshi Ren is a professor in the School of Information and founding director of the Center for Human-Engaged Computing (CHEC) at Kochi University of Technology. He is founding president and honorary lifetime president of the International Chinese Association of Computer Human Interaction (ICACHI). He was ACM CHI 2015 local hero, a visiting professor at the University of Toronto, a visiting faculty researcher at IBM Research (Almaden), and visiting guest/chair professor at several universities in China. He is currently adjunct professor, PhD student supervisor of University of Science and Technology Beijing. He is a senior member of the ACM, a senior member of the IEEE, a member of the IPSJ, the IEICE, and the Human Interface Society. Professor Ren has been working on fundamental studies in the field of human-computer Interaction (HCI) for over 25 years. His research interests include all aspects of HCI, particularly human performance models, pen-based interaction, multitouch interaction, eye-based interaction, haptic interaction, gesture input, game interaction, user interfaces for older users and for blind users. He and his colleagues have established a unique research framework based on information technology, incorporating methodologies such as human performance modeling, developing new algorithms, conducting user studies, and systematically testing and applying HCI theory to applications.

Giuseppe Santucci is associate professor at the Department of Computer Science of Sapienza University, Rome, where he teaches courses on Python, software engineering (in English), and visual analytics (in English). His main research activities concern HCI, information visualization, focusing on quality aspects of information visualization techniques, and the application of visual analytics to predictive scenarios. He published more than 100 papers in international journals and conferences on these topics.

Jinwook Seo is an associate professor in the Department of Computer Science and Engineering at Seoul National University (SNU) in Korea. He founded the HCI laboratory at SNU in 2009. His research interests include HCI, information visualization, and biomedical informatics. He received his BS and MS degrees in computer science from Seoul National University in 1995 and 1997, respectively, and then he taught at Korea Air Force Academy for three years. He later worked with Ben Shneiderman at the HCI Laboratory, University of Maryland, and earned his PhD in computer science in 2005. He received the Long-Term Impact Award at IEEE VisWeek Discovery Exhibition in 2009 for his work on the hierarchical clustering explorer. He has served on various conference organization and program committees such as CHI, IEEE InfoVis, IEEE VAST, and IEEE PacificVis.

Teddy Seyed is currently the first entrepreneurial PhD candidate in computer science at the University of Calgary. His academic research primarily focuses on usability and interaction design within wearables and fashion technology. The entrepreneurial side of Teddy's research focuses on the commercialization of novel wearable devices. His academic and leadership pursuits have also been recognized by a number of top scholarships in Canada and North America. He is also a member of the agile surface engineering (ASE) group at the University of Calgary headed by Dr. Frank Maurer and the Interactions Lab at the University of Calgary, co-led by Dr. Tony Tang.

DongHwa Shin received a BS degree in computer science from the Korea Advanced Institute of Science and Technology (KAIST). He is currently working toward the PhD degree in the Department of Computer Science and Engineering at Seoul National University. His research interests include visualization and visual analytics for high-dimensional data.

Chaklam Silpasuwanchai is a postdoc in the Center for Human-Engaged Computing (CHEC) at Kochi University of Technology. His research interest lies in HCI, particularly in understanding the underlying mechanisms of emotion and cognition and examine their HCI significance in terms of their role to enhance human capacities or user experience. His recent research focused on enhancing human capacities through motion games and meditation. Other recent research focuses on understanding how users learn using cognitive models. He is a recipient of the KAKENHI grant (Ministry of Education, Japan) to work on a project titled Investigating Cognitive Enhancement of Elderly People by Using Motion Video Games. He is also a member of a JST-Finland grant (with Aalto University User Interfaces Group) to work on a project titled User Interface Design for the Ageing Population. He currently serves as vice-secretary general in the International Chinese Association of Computer Human Interaction (ICACHI) and recently served as general co-chair in the International Symposium on Interactive Technology and Ageing Populations 2016.

S. Shyam Sundar (PhD, Stanford University 1995) is a Distinguished Professor and the founding director of the Media Effects Research Laboratory at Penn State University's College of Communications (http://comm.psu.edu/people/individual/s.-shyam-sundar). His research investigates social and psychological effects of technological elements unique to online communication, ranging from websites to newer social media. He edited the *Handbook of the Psychology of Communication Technology* (Wiley, 2015) and currently serves as editor-in-chief of the *Journal of Computer-Mediated Communication*.

Jinping Wang is a PhD student in the College of Communications at the Pennsylvania State University. She earned a master's degree in communication studies from Renmin University of China in Beijing, 2016. Her primary research interest pertains to the psychological effects of communication technologies. She examines how users' interactions with the technology itself and with other users through the technology influence their perceptions, emotions, and behaviors.

Michel Wedel is the Pepsico Chaired Professor of Consumer Science at the Robert H. Smith School of Business, and a Distinguished University Professor, at the University of Maryland. He has a PhD in Marketing from Wageningen University, 1990, and MSCs in biomathematics and statistics, 1975. His main research interests are in the application of statistical and econometric methods to problems in marketing. Much of his recent work addresses issues in visual marketing, using eye-tracking technology. He has written books on market segmentation and visual marketing. He is area editor of *Marketing Science* and the *Journal of Marketing Research*. Michel has published about 20 book chapters and 175 articles, in such journals as *Marketing Science*, the *Journal of Marketing Research*, the *Journal of Marketing*, the *Journal of Consumer Research*, *Psychometrika*, the *Journal of the American Statistical Association*, and the *Journal of Experimental Psychology*, which have been cited close to 16,000 times. His work has been awarded several prizes. He was ranked the most productive economist in the Netherlands and the most productive market researcher in the world. He received the Muller Award for outstanding contributions to the social sciences from the Royal Dutch Academy of the Sciences, the Charles C. Parlin Award, and the Churchill Award for lifetime contributions to marketing research from the American Marketing Academy, and is a fellow of the American Statistical Association and the Institute for Operations Research and Management Science. The University of Maryland has awarded him the Distinguished Scholar teacher award and named him a Distinguished University Professor.

Sehi L'Yi received the BS degree in computer science and engineering from Chungbuk National University. He is currently working toward a PhD in the Department of Computer Science and Engineering at Seoul National University. His research interests include interactive visualizations for bioinformatics data.

Acknowledgments

We, Kent and Jurek, would like to dedicate this handbook to our loving and devoted wives, Karen and Maìre Doṁnat, respectively, who have long held our sanity and wiped our brows when necessary; to one another who have helped each other to overcome the pitfalls of life and to be academic and spiritual brothers; to our many colleagues in computer science, engineering, psychology, and other disciplines who have worked hard and long to bring us to this point in history; and above all to the Creator of all things visible and invisible, which includes the human-computer interface and beyond.

Such a large and ambitious project would have been impossible without the continued help and dedicated assistance of the wonderful team at Wiley-Blackwell from the editor of the psychology series, Andrew Peart, who first anticipated the need for this handbook to the many who helped to shepherd it through to completion: our project editors in order, Karen Shield, Roshna Mohan, and finally Silvy Achankunji; our copyeditor, David Michael, whose sensitive touch and appreciation of the finer points of APA style were refreshing and educational; our production editor, Kumudhavalli Narasimhan, who brought this project to its completion; and finally, for our cover design, Monica Rogers.

Part VI
Interfaces

22

Multisurface Environments

Teddy Seyed and Frank Maurer

Introduction

Traditionally, users have interacted with information on a single computer with one or more monitors following the keyboard-and-mouse paradigm. This paradigm was reflected in the hardware environment itself remaining relatively consistent. For example, users of word-processing applications could work in either home or office environments without much worry about the changes in the computing technology used or the interaction with the information itself. More recently, there has been a paradigm shift in computing technologies, with newer forms of interaction technologies becoming commonplace. The modern computing environment now includes desktops, laptops, smartphones, tablets, digital tabletops, and high-resolution wall displays—and data interaction can involve several of these at the same time in a sequence of steps.

The notion of a computing environment that consists of a multitude of different devices with different form factors was first envisioned by Mark Weiser (1999). This type of computing environment—a *ubiquitous computing* environment—describes machines in the environment that "come in different sizes, each suited to a particular task" and are "interconnected in a ubiquitous network." Returning to the word-processing example, in such an environment (in either home or office settings) users have several choices to make: where to do their word processing; what device should be used (e.g. a tablet, laptop, mobile); and what type of word processing tasks they would like to accomplish—reading, editing, or document creation. Undoubtedly, with each new tablet, mobile phone, wall display, or other new device that becomes interconnected in these ubiquitous environments, we come closer to the "natural human environment" as described by Weiser.

One major challenge in building the "natural human environment" is that devices require particular information in order to be physically and/or spatially aware of their surroundings and each other. Information such as the location of people, devices, and the orientation and distance of each device relative to other devices, are all important pieces in building a "natural human environment." While much of this information is available individually through external sensors and built-in device sensors, it is their

The Wiley Handbook of Human Computer Interaction, Volume 2, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

combination into the "natural human environment" that has not fully occurred yet. This combination is called a multisurface environment (MSE), which we discuss in this chapter. Next, we define and describe the evolution of MSEs, followed by an examination of the design space they provide.

Defining Multisurface Environments

A more traditional environment that consists of multiple displays can be described as an "interactive computer system with two or more displays that are in the same general space (e.g., the same room) and that are related to one another in some way" (Nacenta, Gutwin, Aliakseyeu, & Subramanian, 2009). It also includes "systems where multiple displays are connected to a single computer and systems where networked computers link their displays" (Nacenta et al., 2009). In this definition, displays are defined as mechanisms of output, and not necessarily mechanisms of input. This is an important and distinguishing aspect of the definition we provide for MSEs, which is the following:

> A multisurface environment is a spatially aware environment where interaction with information is distributed over several different devices, such as digital tabletops, wall displays, tablets and mobile phones, and the space in between. This distribution of interaction means that multiple users in this environment can perform multiple activities with any number or combination of different components.

In this definition, there is a specific focus on devices and sensors that provide different forms of input for interaction with information—such as multitouch gestures and 3D interactions. Interaction in this definition also means that displays and devices in multisurface environments are not simply means of output but are interactive surfaces that are aware of each other and people. This "emphasizes the nature of many of today's interactive walls, tables, Tablet PCs, desktop displays, laptops and PDAs that often can be interacted upon in addition to be merely the visual display" (Shen, Esenther, Forlines, & Ryall, 2006). With MSEs, applications and content are no longer bound to a single device, but instead distributed amongst the devices in the environment and the space itself.

Figure 22.1, highlights an example of an MSE that contains a digital tabletop, wall display, tablets, and a mobile phone. In this environment, users can freely choose a particular display or device for an appropriate task. However, not all tasks are suited for all the devices or displays. For example, a user who needs to examine large quantities of information (i.e. geospatial data), would likely choose a larger interactive surface, such as the digital tabletop or wall display over the much smaller tablet and mobile devices.

Also emphasized in the definition of MSEs is the notion of spatial awareness, meaning that displays and devices are aware of each other's location, the people in the space and other properties (such as orientation). Applying spatial awareness to Figure 22.1, means that a mobile phone can be aware of the location of the wall display, and if a user were to take the mobile phone and point it at the wall display, the wall display itself would be aware that it is being pointed at with a mobile phone. This example highlights how user interactions can be influenced by the spatial layout

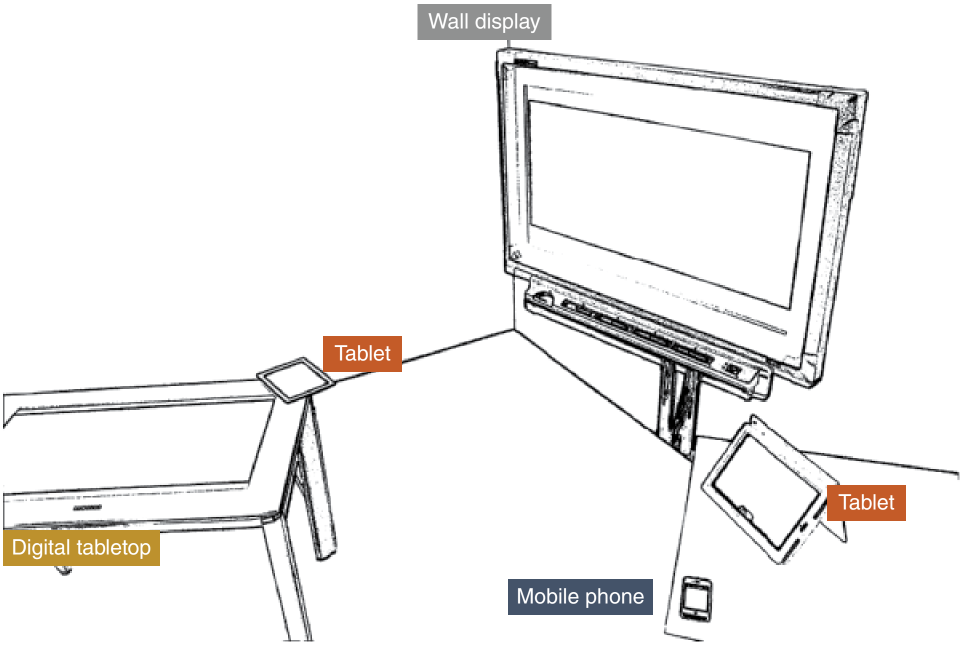


Figure 22.1 An example of a multisurface environment with different displays and devices.

of the environment and is important for several interactions that we discuss in detail later in this chapter.

The Evolution of Multisurface Environments

An early example of an MSE is the Cyworld Control Room created in 1999 (Figure 22.2), which consisted of a large number of individual workstations (each with a single display) and a large wall display that users could interact with using their workstations (http://www.cyworld.com/cymain/?f=cymain). The purpose of this environment was to allow administrators to monitor various aspects of the virtual game world called Cyworld. This type of environment and others in the research literature were found in multiuser meeting rooms or workspaces, where users typically had their own workstations (Stefik et al., 1987).

Another example of an early prototype MSE in this context is Courtyard (Tani, Horita, Yamaashi, Tanikoshi, & Futakawa, 1994), which supported cooperative work by integrating individual workstations with a large shared display that could be manipulated via traditional keyboard and mouse input from individual workstations. The large display acted as a shared workspace where users could place their individual content, prompting collaborative discussions amongst other users in the environment. Prior to these types of early environments, users were typically required to move physically to other individual workstations for collaboration, which wasn't necessarily feasible for large groups of users collaborating (such as city planners). The shared workspace concept was also extended to allow for techniques such as allowing large displays for overview and smaller displays for details. These techniques augmented

Figure 22.2 The Cyworld control room center in Seoul, Korea.

the tasks of the individual fixed displays in several research prototypes, allowing users to use multiple displays simultaneously for a variety of tasks (Cook, Ellis, Graf, Rein, & Smith, 1987).

A key benefit of including multiple heterogeneous screens in workspaces and meeting rooms—displays and large wall displays—is allowing users to divide and organize tasks appropriately and more efficiently, particularly in collaborative tasks (Bandelloni & Paternò, 2004). Much of the early research literature and prototypes focused on *multidisplay* environments, which were designed for collaborative tasks using fixed workstations (or displays) and used mostly keyboards and mice for interactions. A more formal description of a multidisplay environment according to this early literature and research is as follows:

> Definition 1—Multidisplay environments: a multidisplay environment is an environment where multiple displays, such as wall displays and computer monitors can be interacted with by users. Users typically interact with these heterogeneous displays via keyboard and mouse input. Additionally, the displays in the environment are typically in fixed, static locations.

One evident problem with the early prototypes of multidisplay environments is that many interactions for the user were tied to fixed displays and restricted their movements (Bolt, 1984). The restriction of a user's mobility to a workstation and limiting their ability to move themselves according to their tasks can be extremely detrimental in these environments (Luff & Heath, 1998). To address this problem, prototypes began using cameras (moving and fixed) to expand workspaces beyond individual workstations (Heath, Luff, & Sellen, 1995). These cameras were used to combine the individual workstations of users in separate locations into one shared workspace for all users. However, research highlighted that these technologies had a negative effect on how users physically moved in the environment and how they collaborated with other users as they began to move uncomfortably due to the presence of cameras (Gaver, Smets, & Overbeeke, 1995). Ultimately, user mobility plays a vital role in MSEs when collaboration is important.

To address the lack of mobility in multidisplay environments, the use of devices with mobile properties in combination with fixed workstations and wall displays was suggested (Gaver et al., 1995). Multisurface environments support user and device mobility and go beyond mouse / keyboard interactions. Several different prototypes of MSEs that incorporated mobile technologies—such as personal digital assistants (PDAs), handheld tablets, and laptops—began to appear. Rekimoto used the concept of mobility to introduce a prototype multisurface whiteboard application, where users could use a handheld tablet to draw upon and then later transfer their drawings to a shared wall display (Rekimoto, 1998). The handheld served as private workspace for users to create drawings, while the wall display acted as a public space where drawings could be shared. In a similar fashion, the enhanced table environment integrated laptops, a shared wall display, and a physical table (Koike, Nagashima, Nakanishi, & Sato, 2005). Users with laptops had distinct workspaces and could transfer content to the shared displays that were made interactive through enhancements from a projector and a hand-tracking system. It is important to note that in enhanced table, the table was the shared center of collaboration—despite the mobility of laptops in the environment—and this was mirrored in several other prototypes that attempted to address mobility within MSEs (Sugimoto, Hosoi, & Hashizume, 2004). This highlights a similarity to early prototypes in which fixed displays were arranged around a table that indirectly served as a center of collaboration (Elwart-Keys, Halonen, Horton, Kass, & Scott, 1990).

The integration of a digital tabletop into a MSE supports the notion of a "natural human environment" (Weiser, 1999), as users have been found to mirror and extend their existing work practices when using them (Brignull, Izadi, Fitzpatrick, Rogers, & Rodden, 2004). Consequently, this means similar practices are performed by users with both digital and physical tabletops. The i-Land project was an early prototype MSE that incorporated a digital tabletop (Streitz et al., 1999) (Figure 22.3). Workstations that were previously fixed became integrated into chairs, which could be moved anywhere in the environment and could also interact with a large, high-resolution wall display. A digital tabletop served as a collaboration center, where users were able to share and interact (via touch input) with content. A similar technique was used with smaller, more portable digital tabletops, which could be interconnected to other surfaces to form another shared digital large tabletop. The mobility allowed within i-Land and its use of a digital tabletop as a shared space provided motivation for several other research prototypes that followed (Everitt, Ryall, & Forlines, 2006). Additionally, MSEs also began incorporating different methods of input, such as pointing gestures and voice commands, to create more natural MSEs.

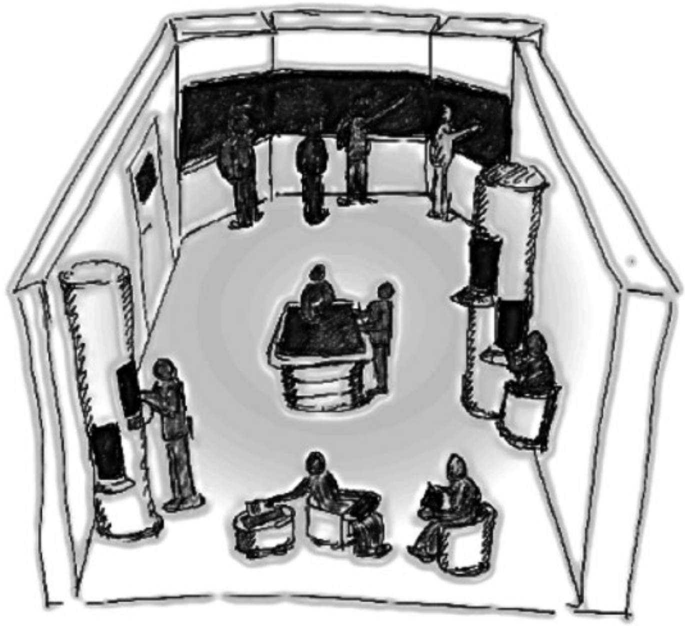


Figure 22.3 The i-Land MSE. *Source:* Streitz et al. (1999).

The trend to use newer mobile technologies—laptops, PDAs, tablets, and digital tabletops—along with different input modalities—keyboard and mouse, touch input, or spatial gestures—highlights the evolution of multidisplay environment to multisurface environments. A definition can be formalized as follows:

> Definition 2—Multisurface environments: a multisurface environment is an environment where multiple displays, some mobile and some fixed—wall displays, laptops, PDAs, tablets, digital tabletops and computer monitors—can be interacted with by users. Users interact with these heterogeneous surfaces through keyboard and mouse, touch input, multitouch gestures or voice commands.

Definition 2 brings two key changes: mobility and variety of input. Displays are no longer in fixed physical locations and interactivity is created from a variety of input sources.

Another important concept in MSEs such as i-Land (Streitz et al., 1999), Co-Mem-iRoom (Fruchter, Saxena, Breidenthal, & Demian, 2007) and others is spatial awareness (Kohtake et al., 2005).

In the context of MSEs, spatial awareness refers to how aware the environment is of the location of its users, displays, and the effect of the location on interactions. In i-Land, no spatial awareness exists as users can move around with displays and no interaction with a display requires knowledge of other displays in the environment (Figure 22.3). For example, users in i-Land can move their content from a chair to the wall by selecting the destination in a graphical menu on the chair itself. This interaction means spatial awareness must be formed in the user's mind before selecting the destination device—the system does not need to understand spatiality. This lack of spatial awareness also means that a user in i-Land is unable to take a portable digital tabletop to the wall display and transfer content based on physical proximity.

In several prototype MSEs, interactions across displays is created with interfaces that have graphical representations of the displays and the environment (a “world in miniature approach”), menus with a list of available displays in the environment (menu-based approach), or novel methods such as “portals,” which allow users to drag content between displays (Wigdor, Shen, Forlines, & Balakrishnan, 2006). A more recent example of an MSE is Microsoft’s SmartGlass application (https://gizmodo.com/5915553/what-is-xbox-smartglass), which allows users to create a connected environment that consists of a gaming console (which is typically connected to a display) and any number of mobile phones or tablets. The tablets and mobile phones can be used to control a display that is connected to the gaming console or vice versa.

Spatial awareness means that all surfaces in a MSE are able to access the absolute physical location of all other surfaces and the users of the space at any time. Limited spatial awareness weakens this—surfaces might know the relative locations of some devices or are able to identify specific devices at certain points in time. For example, in a MSE that has limited spatial information, it is not possible to know the exact physical locations (or orientations) of other devices at all times.

Limited spatial knowledge can be created based on information from sensors such as gyroscopes, and accelerometers, RFID tags, or QR codes. For example, a tablet’s camera can identify a surface using a QR code that is attached to it and use this knowledge to transfer contents to it. Despite the limited spatial information in these MSEs, it has been shown that spatially aware displays provide natural complements or substitutes to direct interaction with other displays in multidisplay environments (Fitzmaurice, 1993). For example, Yee used a sensor-equipped PDA to navigate through custom-designed applications (e.g. a map application) (Yee, 2003). The results of a usability study found that users preferred this method of navigating digital information and that it was significantly more effective compared to other input methods used in the study, such as keyboard and mouse or touch.

One important outcome of spatially aware environments is the resulting changes in how users interact with the displays. A common technique employs mobile devices to control content on a large digital display using simulated spatial awareness (Ballagas, Rohs, & Sheridan, 2005). These techniques included using the mobile display through different physical input mechanisms to transfer content such as sweeping, pointing, and shooting. Similarly, chucking interactions combine touch with accelerometer input on a mobile device, letting users physically toss content to displays in different locations (Hassan, Rahman, Irani, & Graham, 2009). Other proposed additional interactions with mobile devices include throwing and tilting to transfer content to a large digital display (Dachselt & Buchholz, 2009).These types of interactions—interactions with displays or devices instead of on them—are only available due to the spatial awareness of the display or device itself and highlights a change from definition 2, where the only input mechanisms for displays were either keyboard and mouse, touch input or gestures. As a result, a further updated definition is as follows:

> Definition 3—Spatially aware multisurface environments: a spatially aware multisurface environment is an environment where multiple displays, either mobile or fixed—wall displays, laptops, PDAs, tablets and computer monitors—can be interacted with by users using physical location and movement in space. Since some of these devices are spatially

> aware, users can interact with these heterogeneous displays through different input modalities—keyboard and mouse, touch input, body movements, 3D gestures—or with the displays themselves—via intuitive gestures such as throwing or pointing.

A recent example is provided by the Wii U gaming platform by Nintendo, where users have the ability to control content on a screen using a secondary display and shaking gestures (http://www.nintendo.com/wiiu) (Figure 22.4).

Although not all displays are spatially aware in MSEs (as per Definition 3), in order to reach the true "natural human environment" (Weiser, 1999), emphasis should be placed on "human." That is, the users and their movements should have an important role by being recognized. From the perspective of MSEs, this means the environment is spatially aware of all devices and users. Newer, commercially available technologies, such as the Microsoft Kinect sensor (https://developer.microsoft.com/en-us/windows/kinect) and Vicon Motion Capture System (http://www.vicon.com/products/camera-systems), allow for spatial tracking of individuals, and, when combined with sensor information such as orientation and position from the devices themselves, it is possible to create multisurface environments where both users and devices are tracked. These multisurface environments can also consider proxemics, the spatial relationship between users and devices, to enhance interactions (Ballendat, Marquardt, & Greenberg, 2010).

LightSpace is an example of a MSE that utilizes spatial tracking of both users and projected-upon surfaces to allow for intuitive and novel interactions (Wilson & Benko, 2010). Users are able to manipulate content on a tabletop and then transfer it by

Figure 22.4 A "shake" gesture on the handheld controller shaking content on the TV for the Wii U from Nintendo.

touching the content and the target display simultaneously. Alternatively, a user in this environment may also pick-up content with their hands and drop it onto another target display.

The Code Space MSE, built to support collocated, small developer group meetings, used a combination of spatially aware mobile touch devices, the Microsoft Kinect sensor and multitouch screens (Bragdon, DeLine, Hinckley, & Morris, 2011). Users could perform cross-device interactions with a combination of in-air pointing (with or without a device), and touch.

Lastly, proxemics have been explored in MSEs, where users could interact with a touch-enabled wall display that reacted to their distance, identity, physical location, orientation, and even devices such as a mobile phone or tablet (Ballendat et al., 2010). These MSEs and interaction techniques and others highlight the potential of full spatial awareness in multidisplay environments and motivated Definition 3, which places an emphasis on the user and the "natural human environment" (Weiser, 1999). It is also important to note that, as multisurface environments continue to evolve, so will their definition, particularly with the rise of newer technologies such as wearables.

The Interaction Space

In MSEs, where any number of users and devices can exist simultaneously, an inherent problem for interaction designers is how to move content and control across different displays. Movement of content in traditional environments with multiple monitors is typically accomplished with a keyboard and mouse. In a MSE where the interaction is distributed over several different devices, moving content still remains a challenge, as does designing the environment itself. In this section, we discuss a number of unique and effective approaches that have been established in the literature for MSE interactions, which can generally be categorized into approaches that use graphical user interfaces (GUI), tangibles or objects (physical), gestures or proxemics. We then discuss the important factors of designing the output for an MSE.

Graphical user interfaces (GUIs)

In many of the early MSEs described in the previous section, a GUI was the most common approach used to transfer content between displays. The GUI approach, regardless of the technology used in a MSE, has traditionally been the most familiar interface to users. The Windows, Icons, Menu, and Pointer (WIMP) GUI paradigm, which began with Douglas Engelbart during the mid-1960s and is now common to modern operating systems (Engelbart & English, 1968), is a well-established paradigm and components within WIMP can be utilized in a number of different GUI approaches for MSEs.

Traditional menus A menu in a GUI provides a list of options or commands for an application or computer system. In an MSE, a similar approach can be used to transfer content between devices and displays. The TeamSpace MSE comprises several laptops and a shared wall display, and utilizes a menu-based GUI approach to transfer content in a collaborative environment (Hebert & Chen, 2005). Users are able to select other devices and displays from a menu and send content to the selected device or display.

Users are also able to take control of the shared wall display through a menu. The menu-based approach is also used for driving interaction in a geospatial applications designed for MSEs. Menus on a tablet are used to control information (such as layers and drawings) that appears on a digital tabletop and wall display (Forlines, Esenther, Shen, Wigdor, & Ryall, 2006). The menu shows what information a device or display is currently displaying and users can simply choose to add or remove information for a particular display by selecting in the menu. Alternatively, Dynamo uses a menu and icon-based GUI approach to let users transfer content (Izadi, Brignull, Rodden, Rogers, & Underwood, 2003). The icons represent the individual displays in the environment and, in the menu, users are able to drag content (webpages, pictures, videos) to the appropriate targeted device icon.

In general, GUI approaches for multidisplay environments list devices and displays in the environment in a text-based menu or they use icons in a menu to represent the devices. Clicking or dragging is then used as a means of interaction to facilitate the transfer of content. This type of approach is extremely visible in the consumer space, with technologies such as Apple's Airplay letting users share music and video content through various Apple products.

World in miniature The world-in-miniature (WIM) approach for GUIs maps a physical interaction space to a 2D GUI. These types of interfaces are typically presented in "top-down" views with icons or text spatially mapped to specific devices or users in an environment. This type of spatial mapping allows users to understand easily spatial relationships of other users and devices in the environment. Consequently, it is significantly faster for users to send content to a targeted device compared to a menu-based approach (Gostner, Rukzio, & Gellersen, 2008). This is reinforced by prior MSE research indicating that users are inclined to think of spatial relationships with devices either in terms of themselves or the environment, similar to how they would with real-world objects (such as a cup or a book) (Ha, Wallace, Ziola, & Inkpen, 2006). However, the WIM approach may not be faster in cases where the number of devices is either extremely low (e.g. 1 or 2 devices) or high enough to clutter an interface. One alternative for WIM approaches for multidisplay environment allows users to select a display from a spatially mapped WIM GUI and then select content or commands to send to a targeted display (Figure 22.5) (Biehl & Bailey, 2006). Another approach to WIM, where novel interface elements are used to provide links between different displays, is to allow users to drag to conduits (called cords) that are mapped to different displays or to match colors and shapes of the different displays (Wigdor et al., 2006).

Physical interactions

The interactions that users have without technology can primarily be considered forms of physical interactions and they may include physical objects, e.g. swinging a golf club or throwing a ball. The mapping of these physical interactions to the available devices and users in a multidisplay environment results in interactions that map conceptually to real-world interactions. The techniques in which devices are used in physical approaches can differ greatly, however.

Physical objects as a medium Utilizing a physical object as a means of transferring content in a multidisplay environment is a useful technique for designers of MSE interactions because, typically, this type of interaction maps to a concepts with which

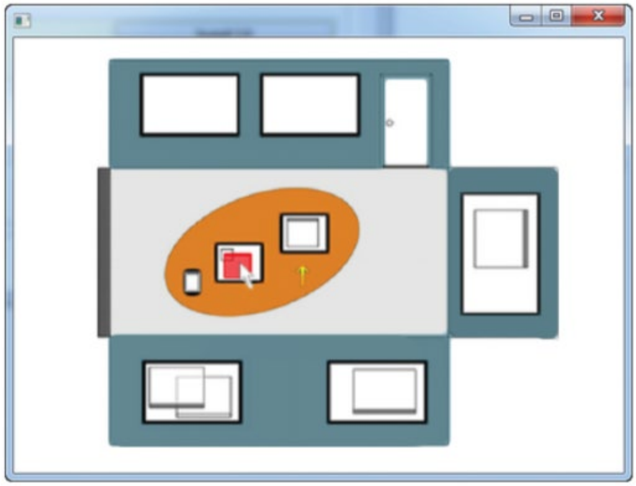


Figure 22.5 A WIM approach to selecting displays. *Source:* Biehl and Bailey (2006).

(a) (b)

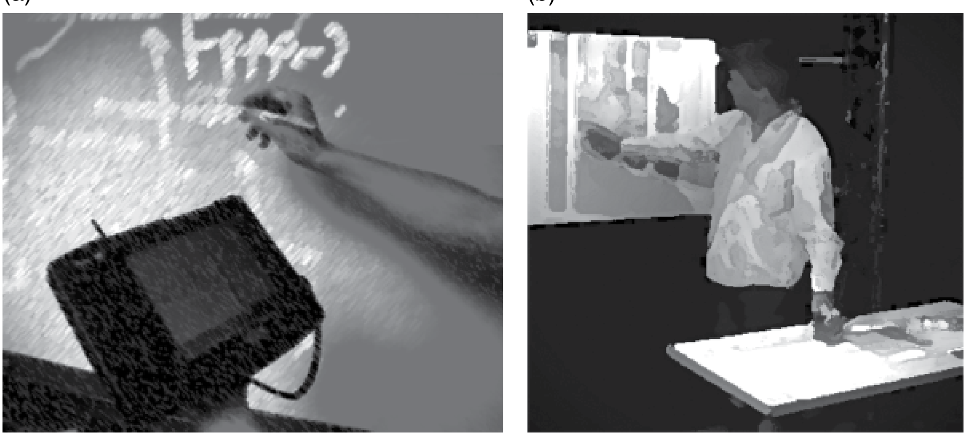


Figure 22.6 Interaction Examples. (a) Pick and drop interaction with a pen (Rekimoto & Nagao, 1995); (b) Direct contact interaction between a tabletop and a wall display (Wilson & Benko, 2010).

users are familiar. For example, a pick-and-drop interaction technique could allows users to select (or "pick") content on one device by tapping it with a pen and place (or "drop") it on another device by tapping it again with the same pen (Rekimoto & Nagao, 1995). The pen in this case provides a physical medium to transfer the information (as shown in Figure 22.6a). Another example of using a physical medium as an interaction approach is the previously mentioned i-Land, where users were able to link information to any arbitrary object (a key, watch, or block of wood, for example), which could then be placed onto digital tabletops that identify and transfer the data linked to the object (Streitz et al., 1999).

Direct contact For interactions in MSEs, direct contact implies directly contacting one or more devices to transfer content or control. An example of this direct-contact interaction approach is provided in LightSpace, in which users were able simultaneously to contact a wall display and a tabletop to transfer content (Wilson & Benko, 2010). A user becomes a conduit by directly contact the two displays simultaneously, as shown in Figure 22.6b. Another technique is illustrated in Augmented Surfaces, where users are able to contact devices directly in order to transfer content (Rekimoto & Saitoh, 1999).

Gestures and proxemics

Gestures provide additional benefits for interactions in MSEs, particularly because they can be combined with touch-based displays (Krahnstoever, Kettebekov, Yeasin, & Sharma, 2002). Proxemics is a sociological field that examines personal space and social interactions of people, and is defined by physical interaction zones based on distance (Hall, 1968). Proxemic interactions are based on this field and are more concerned with the factors that affect spatial relationship between devices, such as position, orientation, and movement (Ballendat et al., 2010). Combining the gestural approach and proxemic techniques results in a powerful mechanism of transferring content, as both the gestures and devices are now spatially aware of the environment.

Gestures on devices Interacting directly on a display (typically via touch input) in an MSE is an approach that users can use to select and transfer content. For example, in Code Space, users are able to flick content from a mobile display to wall displays or other users with mobile displays (Bragdon et al., 2011). Users are also able to retrieve content from other displays by performing a downward swiping gesture. A reverse interaction for sending content is also supplied, where users are able to select content with a mobile display and then perform an upward swiping gesture to send content to the wall display. An example of a flick gesture made on a tablet to send content to a digital tabletop is shown in Figure 22.7.

Gestures with devices In MSEs, gestures with devices typically involve the physical movement of devices. The rotation of a device was one of the gestures first

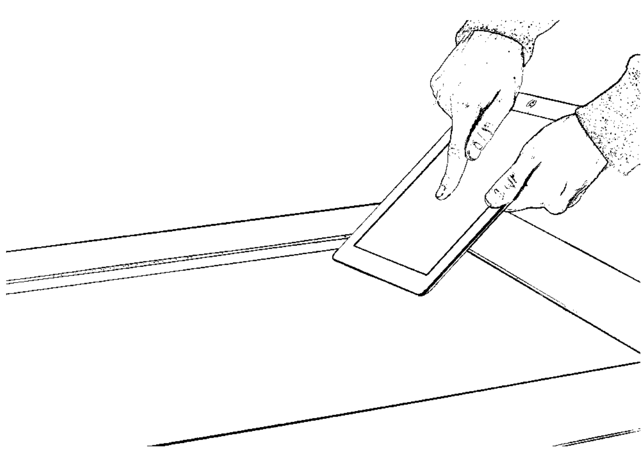


Figure 22.7 Flick gesture.

explored in this space (Bhandari & Lim, 2008). To trigger this gesture, a user rotated a mobile phone from a horizontal to vertical position. Other gestures explored included the shaking gesture, where a mobile device is rocked back and forth along an axis (Bhandari & Lim, 2008). This work was limited, however, as it triggered actions on the device and did not effect the environment. Subsequent work linked these gestures to MSEs and added gestures such as "throw" from a mobile device to a digital tabletop, as well as "pulling" from a digital tabletop to a mobile device (Döring, Shirazi, & Schmidt, 2010). The "throw" gesture and its subsequent iterations in the literature is a canonical example of using a device as a gesture mechanism (Dachselt & Buchholz, 2009).

Gestures without devices Gestures without devices in a MSEs are typically performed with fingers, hands, and arms, and can be further augmented with techniques such as voice commands and eye tracking (Bolt, 1984). Another example of this technique is also provided by LightSpace, where the positional tracking built into the system and roof-mounted projectors, allows users to carry a digital object in their hand from one location in the environment to another (Wilson & Benko, 2010). Pointing is an another common interaction technique that doesn't require devices, as shown in Code Space (Bragdon et al., 2011). Table 22.1 summarizes the systems described in the interaction space section.

Designing the Environment

In MSEs, the spatial layout of the environment and the inherent spatial separation between static displays, mobile displays, and users is extremely important at an interaction level (Su & Bailey, 2005). This separation is defined as displayless space (the space between devices or displays) (Nacenta et al., 2009). How it is regarded in the

Table 22.1 Summary of systems described in interaction space.

Interaction space	GUIs	Traditional menus	Hebert and Chen (2005) Forlines et al. (2006) Izadi et al. (2003)
		World in miniature	Gostner et al. (2008) Ha et al. (2006) Biehl and Bailey (2006) Wigdor et al. (2006)
	Physical interactions	Physical objects as a medium	Rekimoto & Nagao (1995) Streitz et al. (1999)
		Direct contact	Wilson and Benko (2010) Rekimoto and Saitoh (1999)
	Gestures and proxemics	Gestures on devices	Bragdon et al. (2011)
		Gestures with devices	Döring et al. (2010) Dachselt and Buchholz (2009)
		Gestures without devices	Bolt (1984) Wilson and Benko (2010) Bragdon et al. (2011)

spatial layout of an environment is a constraint in the design of the environment. Devices (or displays) in an MSE could be linked together in a common space (also known as spatially continuous) considered separate (also known as spatially distinct) or a combination of both (a hybrid).

Continuous

In traditional PC-based environments where multiple displays can exist in a variety of physical arrangements, it is common for users to link the displays into a single and continuous display space. Applications can then map a continuous information space onto this continuous physical space. The continuous physical space allows interaction to be across displays, as a user can move a mouse cursor continuously in the direction of any destination display. The displays are typically mapped to a conceptual physical location and several modern operating systems utilize this technique when linking multiple displays (Figure 22.8a). In addition, software solutions also exist to allow PCs, laptops and other types of displays to create a continuous display space (Microsoft Garage Mouse without Borders, https://www.microsoft.com/en-ca/download/details.aspx?id=35460).

A similar technique was applied to mobile displays, where users were able to attach display panels to other display panels physically, creating shared continuous spaces (Kohtake et al., 2005). An advantage of this particular system was that it enabled users to spatially configure the mobile displays for different tasks (Hutchings, Smith, Meyers, Czerwinski, & Robertson, 2004). This concept was built upon in spatially continuous interactive workspaces that incorporated laptops, wall displays, and a desktop (Figure 22.8b) (Rekimoto & Saitoh, 1999). These interactive workspaces mapped each of the individual components into a shared space that represented their associated physical location. Users were able to move content between displays by simply dragging it to the next display that was spatially continuous (Hinckley, Ramos, Guimbretiere, Baudisch, & Smith, 2004).

To show a continuous information space across several physically separate devices in an MSE correctly, the empty space between the devices needs to be treated adequately. Figure 22.9 highlights an MSE that shows a single, continuous information space, in this case: a map. The different displays show different parts of a large map, highlighting the spatial continuity as well as the physicality of the display space.

Projecting surfaces (Pratte, Seyed, & Maurer, 2014) or approaches like RoomAlive (Jones et al., 2014) can be used to fill the empty space between displays with output showing the continuity of the information space.

Discrete

Multiple displays can also show different information spaces, in a way of using physicality to highlight different kinds of information. An example of this would be a workspace containing a laptop, mobile phone, and a desktop with a large display running a media application. Both the laptop and mobile phone each have a different means of controlling the media application on the large screen, albeit each device contains different content, which highlights the nature of discrete displays that are related.

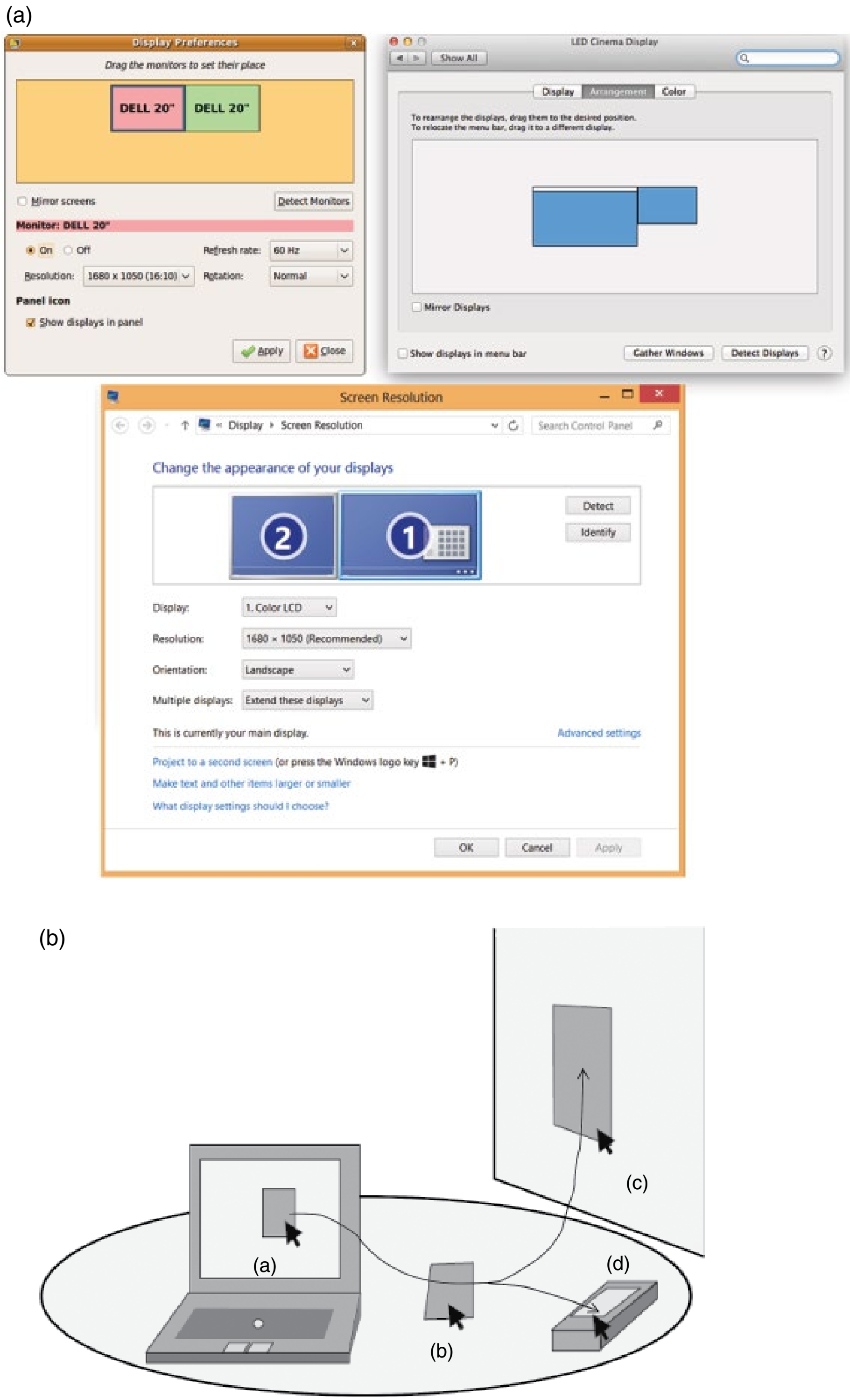


Figure 22.8 Examples of workspaces. (a) Continuous desktop configurations in a variety of operating systems; (b) A spatially continuous workspace that allows a user to drag to different devices and objects. *Source:* Rekimoto & Saitoh (1999).

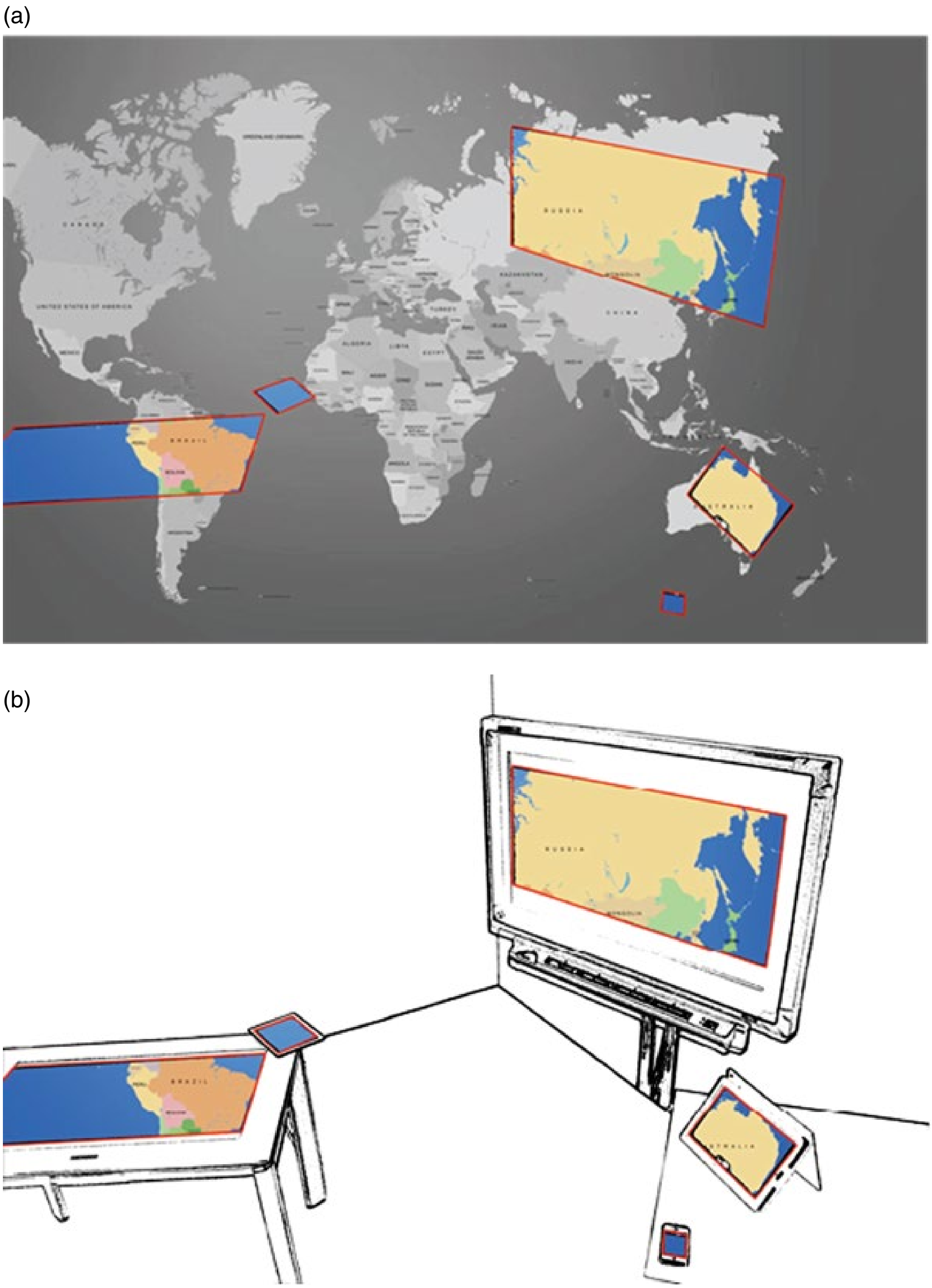


Figure 22.9 (a) A shared map that is physically mapped and linked to device in a spatially continuous MSE. If a device is physically moved in (b), the device displays the spatially connected map location.

Figure 22.10 A spatially discrete MSE. Each device displays a different map, highlight the notion of each being discrete in the environment.

Including i-Land (Streitz et al., 1999), another example of a spatially discrete workspace is one where a digital tabletop is used to facilitate collaborative situation assessment and decision making for the New York Police Department (Wigdor et al., 2006). In this workspace, two wall displays and a digital tabletop are provided—each with distinct digital workspaces that have different menus and content. Although the displays are spatially discontinuous, users are able to share content between the displays through icons that represent each display (a WIM approach).

Figure 22.10 highlights a spatially discrete multisurface environment where different types of maps are displayed on different displays. Interactions that are possible in this type of multisurface environment include picking and dropping, flicking, pouring, and throwing different maps to different displays.

Hybrid

A hybrid MSE is a combination of spatially discrete and spatially continuous displays. As shown in Figure 22.11, spatially discrete displays showing different maps are combined with spatially continuous displays showing different parts of a shared map.

The WILD Room is an MSE that combines spatially discrete displays and spatially continuous displays into a unique interactive environment (Beaudouin-Lafon et al., 2012). It consists of an ultra-high resolution wall display, digital tabletop, and numerous mobile devices. This hybrid MSE allows for users to have two distinct workspaces in the environment: a spatially discrete private workspace where individual users can view and manipulate content and a spatially continuous shared workspace where users can share content and collaborate. Hybrid environments can provide users private and public workspaces (Lee et al., 2011).

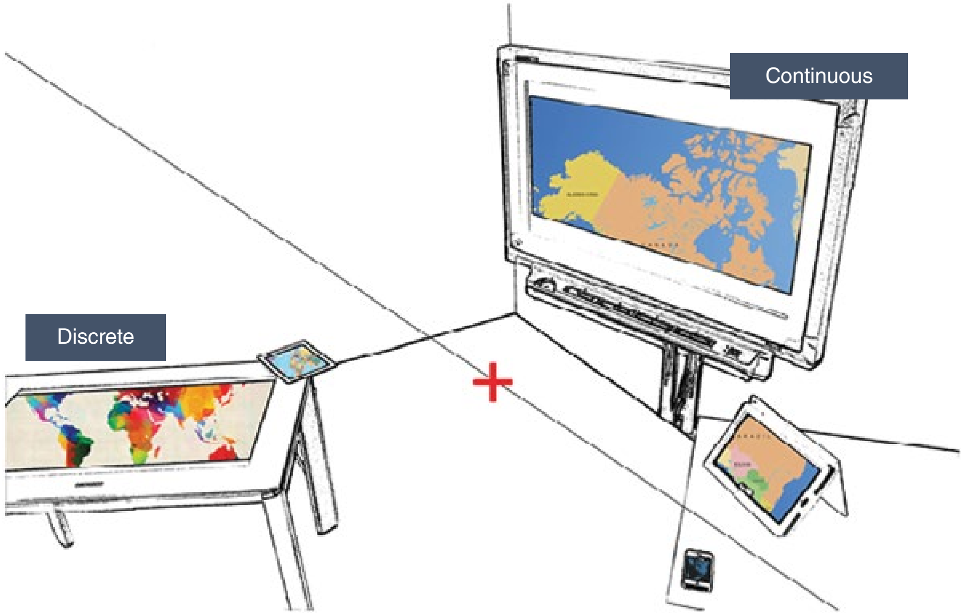


Figure 22.11 A hybrid MSE. Part of the environment contains discrete devices, while others are in a continuous space.

Table 22.2 Summary of systems described in "designing the environment."

Designing the environment	Continuous	Microsoft Garage Mouse without Borders, https://www.microsoft.com/en-ca/download/details.aspx?id=35460
		Kohtake et al. (2005)
		Hutchings et al. (2004)
		Rekimoto and Saitoh (1999)
		Hinckley et al. (2004)
	Discrete	Streitz et al. (1999)
		Wigdor et al. (2006)
	Hybrid	Beaudouin-Lafon et al. (2012)
		Lee et al. (2011)

Ultimately, the choice of the environment for an MSE is dependent upon the nature and mobility of the displays, the spatial awareness provided and the types of tasks users are expected to perform in the environment. Table 22.2 summarizes the systems described in the "designing the environment" section.

Conclusion

In this chapter we presented an exploration into multisurface environments by first defining them, and then discussing their evolution with an emphasis on the changes as displays evolved from stationary to mobile as well as the increase in availability of spatial tracking systems.

Multisurface environments are becoming increasingly commonplace, and with the increase in new types of displays, such as wearables and other novel form factors, the role of MSEs will continue to evolve and change. Nearly all technology tradeshows today present technologies that can either be integrated into or be considered parts of multisurface environments, much like the current generation of gaming consoles. Furthermore, the inherent potential these environments contain is now being realized in a wide variety of fields—such as medicine, oil and gas, and emergency management (Chokshi, Seyed, Marinho Rodrigues, & Maurer, 2014).

References

Ballagas, R., Rohs, M., & Sheridan, J. G. (2005). Sweep and point and shoot. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems (CHI '05)* (p. 1200). doi:10.1145/1056808.1056876

Ballendat, T., Marquardt, N., & Greenberg, S. (2010). Proxemic interaction. In *ACM International Conference on Interactive Tabletops and Surfaces (ITS '10)* (p. 121). doi:10.1145/1936652.1936676

Bandelloni, R., & Paternò, F. (2004). Flexible interface migration. In *Proceedings of the Ninth International Conference on Intelligent User Interface (IUI '04)* (p. 148). doi:10.1145/964442.964470

Beaudouin-Lafon, M., Huot, S., Nancel, M., Mackay, W., Pietriga, E., Primet, R.,…Klokmose, C. (2012). Multisurface interaction in the WILD room. *Computer*, *45*(4), 48–56. doi:10.1109/MC.2012.110

Bhandari, S., & Lim, Y.-K. (2008). Exploring gestural mode of interaction with mobile phones. In *Proceeding of the 26th Annual CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI '08)* (p. 2979). doi:10.1145/1358628.1358794

Biehl, J. T., & Bailey, B. P. (2006). Improving interfaces for managing applications in multiple-device environments. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '06)* (p. 35). doi:10.1145/1133265.1133273

Bolt, R. A. (1984). *Human interface: Where people and computers meet.* New York, NY: John Wiley & Sons, Inc.

Bragdon, A., DeLine, R., Hinckley, K., & Morris, M. R. (2011). Code space. In *Proceedings of the ACM International Conference on Interactive Tabletops and Surfaces (ITS '11)* (p. 212). doi:10.1145/2076354.2076393

Brignull, H., Izadi, S., Fitzpatrick, G., Rogers, Y., & Rodden, T. (2004). The introduction of a shared interactive surface into a communal space. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work (CSCW '04)* (p. 49). doi:10.1145/1031607.1031616

Chokshi, A., Seyed, T., Marinho Rodrigues, F., & Maurer, F. (2014). ePlan Multi-Surface. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces (ITS '14)* (pp. 219–228). doi:10.1145/2669485.2669520

Cook, P., Ellis, C., Graf, M., Rein, G., & Smith, T. (1987). Project Nick: Meetings augmentation and analysis. *ACM Transactions on Information Systems*, *5*(2), 132–146. doi:10.1145/27636.27638

Dachselt, R., & Buchholz, R. (2009). Natural throw and tilt interaction between mobile phones and distant displays. In *Proceedings of the 27th International Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '09)* (p. 3253). doi:10.1145/1520340.1520467

Döring, T., Shirazi, A. S., & Schmidt, A. (2010). Exploring gesture-based interaction techniques in multi-display environments with mobile phones and a multi-touch table. In *Proceedings of the International Conference on Advanced Visual Interfaces (AVI '10)* (p. 419). doi:10.1145/1842993.1843097

Elwart-Keys, M., Halonen, D., Horton, M., Kass, R., & Scott, P. (1990). User interface requirements for face to face groupware. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Empowering People (CHI '90)* (pp. 295–301). doi:10.1145/97243.97299

Engelbart, D. C., & English, W. K. (1968). A research center for augmenting human intellect. In *Proceedings of the December 9–11, 1968, Fall Joint Computer Conference (AFIPS '68) (Fall, part I)* (p. 395). doi:10.1145/1476589.1476645

Everitt, K., Ryall, K., & Forlines, C. (2006). MultiSpace: Enabling electronic document micromobility in table-centric, multi-device environments. In *Proceedings of the First IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP '06)* (pp. 27–34). doi:10.1109/TABLETOP.2006.23

Fitzmaurice, G. W. (1993). Situated information spaces and spatially aware palmtop computers. *Communications of the ACM*, *36*(7), 39–49. doi:10.1145/159544.159566

Forlines, C., Esenther, A., Shen, C., Wigdor, D., & Ryall, K. (2006). Multi-user, multi-display interaction with a single-user, single-display geospatial application. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology (UIST '06)* (p. 273). doi:10.1145/1166253.1166296

Fruchter, R., Saxena, K., Breidenthal, M., & Demian, P. (2007). Collaborative design exploration in an interactive workspace. *Artificial Intelligence for Engineering Design, Analysis and Manufacturing*, *21*, 279–293. doi:10.1017/S0890060407000285

Gaver, W. W., Smets, G., & Overbeeke, K. (1995). A virtual window on media space. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '95)* (pp. 257–264). doi:10.1145/223904.223937

Gostner, R., Rukzio, E., & Gellersen, H. (2008). Usage of spatial information for selection of co-located devices. In *Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '08)* (p. 427). doi:10.1145/1409240.1409305

Ha, V., Wallace, J., Ziola, R., & Inkpen, K. (2006). My MDE: Configuring virtual workspaces in multi-display environments. In *CHI '06 Extended Abstracts on Human Factors in Computing Systems (CHI EA '06)* (p. 1481). doi:10.1145/1125451.1125723

Hall, E. T. (1968). Proxemics. *Current Anthropologist*, *2*(3), 83–100.

Hassan, N., Rahman, M. M., Irani, P., & Graham, P. (2009). Chucking: A one-handed document sharing technique. In T. Gross, J. Gulliksen, P. Kotzé, L. Oestreicher, P. Palanque, R. O. Prates, & M. Winckler (Eds.), *Proceedings of the 12th IFIP TC 13 International Conference on Human-Computer Interaction: Part II* (Vol. 5727, pp. 264–278). doi:10.1007/978-3-642-03658-3

Heath, C., Luff, P., & Sellen, A. (1995). Reconsidering the virtual workplace: Flexible support for collaborative activity. In *Proceedings of the Fourth European Conference on Computer-Supported Cooperative Work (ECSCW)* (pp. 83–89). doi:10.1.1.74.9404

Hebert, A., & Chen, A. (2005). A new collaborative software package. In *Proceedings of the 33rd Annual ACM SIGUCCS Conference on User Services (SIGUCCS '05)* (pp. 109–112). doi:10.1145/1099435.1099460

Hinckley, K., Ramos, G., Guimbretiere, F., Baudisch, P., & Smith, M. (2004). Stitching: Pen gestures that span multiple displays. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '04)* (p. 23). doi:10.1145/989863.989866

Hutchings, D. R., Smith, G., Meyers, B., Czerwinski, M., & Robertson, G. (2004). Display space usage and window management operation comparisons between single monitor and multiple monitor users. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '04)* (p. 32). doi:10.1145/989863.989867

Izadi, S., Brignull, H., Rodden, T., Rogers, Y., & Underwood, M. (2003). Dynamo: A public interactive surface supporting the cooperative sharing and exchange of media. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology (UIST '03)* (pp. 159–168). doi:10.1145/964696.964714

Jones, B., Shapira, L., Sodhi, R., Murdock, M., Mehra, R., Benko, H.,…Raghuvanshi, N. (2014). RoomAlive: magical experiences enabled by scalable, adaptive projector-camera units. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology (UIST '14)* (pp. 637–644). doi:10.1145/2642918.2647383

Kohtake, N., Ohsawa, R., Yonezawa, T., Matsukura, Y., Iwai, M., Takashio, K., & Tokuda, H. (2005). u-Texture: self-organizable universal panels for creating smart surroundings. In M. Beigl, S. Intille, J. Rekimoto, & H. Tokuda (Eds.), *Proceedings of the 7th International Conference on Ubiquitous Computing* (Vol. 3660, pp. 19–36). doi:10.1007/11551201

Koike, H., Nagashima, S., Nakanishi, Y., & Sato, Y. (2005). EnhancedTable: An augmented table system for supporting face-to-face meeting in ubiquitous environment. In H. Murakami, H. Nakashima, H. Tokuda, & M. Yasumura (Eds.), *Ubiquitous computing systems (SE—10)* (Vol. 3598, pp. 117–130). doi.org/10.1007/11526858_10

Krahnstoever, N., Kettebekov, S., Yeasin, M., & Sharma, R. (2002). A real-time framework for natural multimodal interaction with large screen displays. In *Proceedings of the Fourth IEEE International Conference on Multimodal Interfaces* (pp. 349–354). doi:10.1109/ICMI.2002.1167020

Lee, J. Y., Kim, M. S., Seo, D. W., Lee, C.-W., Kim, J. S., & Lee, S. M. (2011). Dual interactions between multi-display and smartphone for collaborative design and sharing. In *2011 IEEE Virtual Reality Conference* (pp. 221–222). doi:10.1109/VR.2011.5759478

Luff, P., & Heath, C. (1998). Mobility in collaboration. In *Proceedings of the 1998 ACM Conference on Computer Supported Cooperative Work (CSCW '98)* (pp. 305–314). doi:10.1145/289444.289505

Nacenta, M. A., Gutwin, C., Aliakseyeu, D., & Subramanian, S. (2009). There and back again: Cross-display object movement in multi-display environments. *Human–Computer Interaction*, *24*(1–2), 170–229. doi:10.1080/07370020902819882

Pratte, S., Seyed, T., & Maurer, F. (2014). Exploring multi-surface interactions in retail environments. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces (ITS '14)* (pp. 393–398). doi.org/10.1145/2669485.2669552

Rekimoto, J. (1998). A multiple device approach for supporting whiteboard-based interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '98)* (pp. 344–351). doi.org/10.1145/274644.274692

Rekimoto, J., & Nagao, K. (1995). The world through the computer. In *Proceedings of the 8th Annual ACM Symposium on User Interface and Software Technology (UIST '95)* (pp. 29–36). doi:10.1145/215585.215639

Rekimoto, J., & Saitoh, M. (1999). Augmented surfaces: a spatially continuous work space for hybrid computing environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems the CHI is the Limit (CHI '99)* (pp. 378–385). doi:10.1145/302979.303113

Shen, C., Esenther, A., Forlines, C., & Ryall, K. (2006). Three modes of multi-surface interaction and visualization. In *Information Visualization and Interaction Techniques for Collaboration across Multiple Displays Workshop associated with CHI'06 International Conference,*. Montreal, QC, Canada. Retrieved from http://www.merl.com/publications/docs/TR2006-025.pdf

Stefik, M., Foster, G., Bobrow, D. G., Kahn, K., Lanning, S., & Suchman, L. (1987). Beyond the chalkboard: computer support for collaboration and problem solving in meetings. *Communications of the ACM*, *30*(1), 32–47. doi:10.1145/7885.7887

Streitz, N. A., Geißler, J., Holmer, T., Konomi, S., Müller-Tomfelde, C., Reischl, W.,… Steinmetz, R. (1999). i-LAND. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems the CHI is the Limit—(CHI '99)* (pp. 120–127). doi:10.1145/302979.303010

Su, R. E., & Bailey, B. P. (2005). Human-computer interaction—INTERACT 2005. In M. F. Costabile & F. Paternò (Eds.), *INTERACT'05 Proceedings of the 2005 IFIP TC13 International Conference on Human-Computer Interaction* (Vol. *3585*, pp. 337–349). doi:10.1007/11555261

Sugimoto, M., Hosoi, K., & Hashizume, H. (2004). Caretta. In *Proceedings of the 2004 Conference on Human Factors in Computing Systems (CHI '04)* (pp. 41–48). doi:10.1145/985692.985698

Tani, M., Horita, M., Yamaashi, K., Tanikoshi, K., & Futakawa, M. (1994). Courtyard. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Celebrating Interdependence (CHI '94)* (pp. 44–50). doi:10.1145/191666.191690

Weiser, M. (1999). The computer for the 21st century. *ACM SIGMOBILE Mobile Computing and Communications Review*, *3*(3), 3–11. doi:10.1145/329124.329126

Wigdor, D., Shen, C., Forlines, C., & Balakrishnan, R. (2006). Table-centric interactive spaces for real-time collaboration. In *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI '06)* (p. 103). doi:10.1145/1133265.1133286

Wilson, A. D., & Benko, H. (2010). Combining multiple depth cameras and projectors for interactions on, above and between surfaces. In *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology (UIST '10)* (p. 273). doi:10.1145/1866029.1866073

Yee, K.-P. (2003). Peephole displays. In *Proceedings of the Conference on Human Factors in Computing Systems (CHI '03)* (p. 1). doi:10.1145/642611.642613

23

A Natural Language Interface for Mobile Devices

Boris Katz, Gary Borchardt, Sue Felshin, and Federico Mora

Introduction

Humans express their needs almost effortlessly in natural language, and for this reason, constructing machines that can reliably respond to natural language requests has been a longstanding and significant goal in the design of intelligent systems. Early successes—for example, the LUNAR system of Woods, Kaplan, and Nash-Webber (1972), the LADDER system of Hendrix, Sacerdoti, Sagalowicz, and Slocum (1978), and the PLANES system of Waltz (1978)—relied on the use of application-specific grammars to encode various constraints of particular domains or datasets. While this approach enabled these systems to respond to a range of requests in their targeted domains, the systems could not easily be ported to work in related or expanded domains, and they had limited coverage of vocabulary and syntactic constructions used in expressing requests. More recently developed systems—such as those described in Harabagiu, Maiorano, and Pasca (2003); Katz (1997); Nyberg, Burger, Mardis, and Ferrucci (2004), and Weischedel, Xu, and Licuanan (2004)—address multiple domains and employ general-purpose grammars or statistical language interpretation to handle the variety of requests and phrasings of requests that might be submitted by users. With more and more information sources available in networked contexts, plus mobile devices acting as sources and targets of requests, recent systems have adopted a more distributed architecture.

In this chapter, we first discuss some of the primary issues related to the design and construction of natural language interfaces, and in particular, interfaces to mobile devices. Then, we describe two systems in this space: the START information access system and the StartMobile natural language interface to mobile devices. Finally, we discuss recently deployed commercial systems and future directions.

Goals of Natural Language Interfaces

The primary goal of a natural language interface is, of course, to produce appropriate responses to user requests. The requests may ask the interface for information, or they may ask for actions to be performed. The interface may issue any number of responses

The Wiley Handbook of Human Computer Interaction, Volume 2, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

to a single request: it may return several segments of information, for example, or perform several actions.

In general, the appropriateness of an interface's responses can be assessed using variants of the metrics of *recall* and *precision* from the field of information retrieval. In this context, recall can be calculated as the average, over many requests, of the fraction formed by dividing the number of appropriate responses returned by the interface by the total number of appropriate responses possible. Precision can be calculated as the average, over many requests, of the fraction formed by dividing the number of appropriate responses returned by the interface by the total number of responses returned by the interface. Recall is often difficult to compute, as there may be no way to enumerate the full set of possible, appropriate responses to a request—the full set of information segments that might be returned, for example, or all appropriate actions that might be performed in response to a request. Recall can be important but in many cases it is of lesser importance, as in those contexts where the user only needs one or a few satisfying responses. Precision, on the other hand, is almost always important. Information requests that return many inappropriate responses distract the user, waste display space, and may even mislead the user. Inappropriate actions can have serious consequences and must be avoided at all costs.

Another goal of natural language interfaces is domain coverage. Limited-domain systems can be useful in some cases, such as an interface to a single, specific database. In general, however, the recent trend is toward construction of interfaces that provide a single point of interaction with respect to a large quantity of datasets, domains and possible actions.

Ease of use is another important aspect of natural language interfaces. An interface should accept unrestricted natural language input to whatever extent is possible, it should exhibit interactivity for clarification of requests and other purposes, it should respond with multimedia information, it should provide explanations of its behavior when demanded, it should provide a history mechanism for review of past responses, and so forth.

One very important aspect of natural language interfaces is their ability to handle complex requests. The simplest sorts of information requests solicit all available information on a specified topic: e.g., "Tell me about Germany," or some specific property, e.g., "Tell me about Germany's population distribution." More complex requests involve the retrieval of information about relationships between entities, or the execution of commands that involve multiple entities. Even more complex are requests that require the system to perform novel analyses or actions. Finally, some requests may involve a nesting of subrequests to be addressed in combination.

Natural language interfaces to mobile devices must exhibit additional characteristics. These interfaces must be able to respond on the basis of information or actions available not only on the mobile device itself but also on systems linked to the device through network connections. Especially where actions are to be performed, such systems must have extremely high precision, as there are often limited ways to rescind these actions. Coupled with this, mobile devices are often used in environments where the user has divided attention and where excessive system interactivity is not desired; thus the interface must in some cases proactively infer what the user has intended, again with extremely high precision. To accomplish this, a system might need to rely on information gleaned from previous interactions with that individual, elements of the preceding dialog, or other available modalities such as camera input. Interfaces to mobile devices must also be particularly responsive to time and location of the user, and these interfaces must make the best use of limited display space.

START

We first discuss the general-purpose START information access system, which has been in development at MIT since the early 1970s (Katz, 1980, 1990, 1997; Katz, Borchardt, & Felshin, 2006). In its most general question-answering application, START is available as a public server at http://start.mit.edu/. This version of START answers questions in a range of domains including geography, arts and entertainment, history, science, and the very large number of topics and domains covered in Wikipedia. In addition to the general-purpose public START server, several special-purpose servers have been created for specific topic areas, some of which involve the execution of actions in response to user requests and others of which make use of an API to START's language parsing and generation capabilities. Separately, several strategies pioneered by the START system were incorporated into IBM's Watson, which in 2011 defeated the all-time human champions on the quiz show *Jeopardy!* (Hardesty, 2011; Murdock, 2012).

In its traditional question-answering role, START accepts English questions and offers responses that draw on information sources that include structured, semistructured, and unstructured materials. Some of these materials are maintained locally and some are accessed remotely through the Internet. In some cases, responses are calculated dynamically by the system or its allied resources. A particular emphasis of START is that of providing high-precision information access, such that the user may maintain a high degree of confidence that a response, if returned by the system, is appropriate to the submitted question. Figure 23.1 presents a sample request–response interaction with START.

A particularly important aspect of START's design is the use of *ternary expressions* as an internal representation of natural language expressions. Ternary expressions represent language as a set of nested subject–relation–object triples, where the subject and object may themselves be ternary expressions (Katz, 1988; Katz, 1990). The ternary expression representation is a versatile syntax-driven representation of language that highlights significant semantic relations and allows for detailed encoding of syntactic and lexical features. It has proved to be extremely beneficial for START's parsing and question answering capabilities due to its speed, compactness, and accuracy for storing, matching, and retrieving information.

As originally configured during the initial stages of its development, START served to answer English questions on the basis of English statements that had been previously submitted to the system, and this operation underlies much of START's current capabilities as well. When START is presented with an English statement for processing, it parses the statement and encodes it in the form of a set of nested ternary expressions. One can think of the resulting entry in START's knowledge base as a "digested summary" of the syntactic structure of the English sentence. User-submitted questions are then analyzed in the same manner and matched against stored assertions in the knowledge base. Matched assertions are then retrieved and expressed as English responses. (The technique of using natural language annotations, described below, extends this approach and enables START to present additional material and perform computations in response to matches.) Because matching occurs at the level of syntactic structures, linguistically sophisticated machinery such as synonymy, hyponymy, ontologies, and structural transformation rules can all be brought to bear on the matching process, thus achieving capabilities far beyond simple keyword matching.

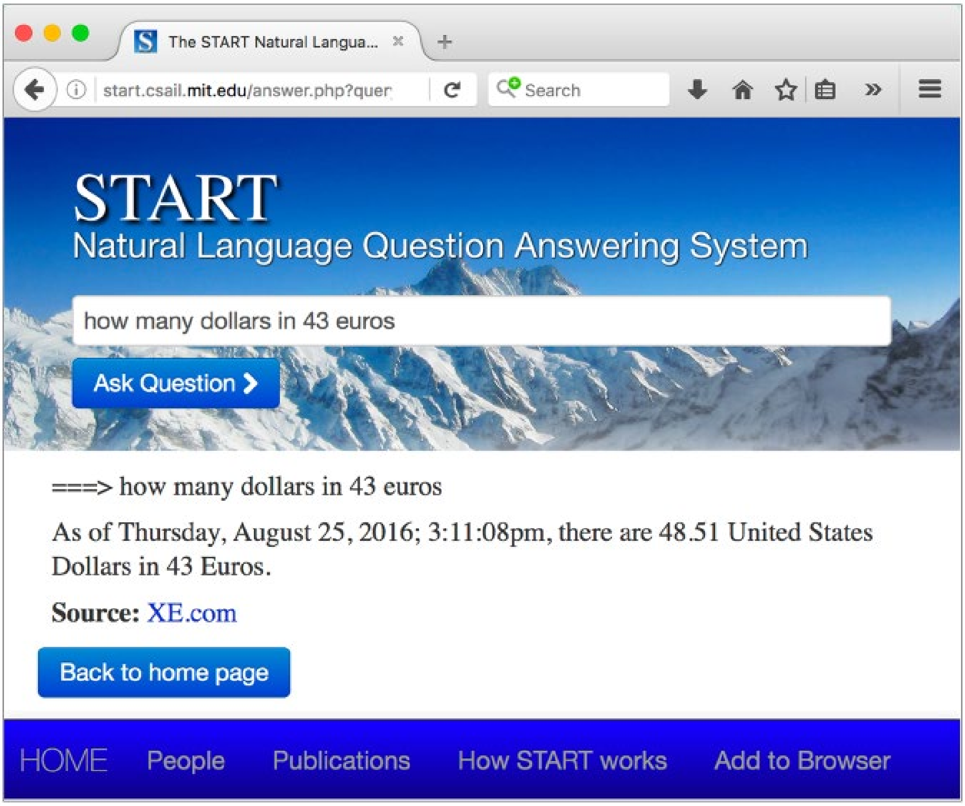


Figure 23.1 START performing a currency conversion.

In particular, structural transformation rules enable the system to find matches despite significant differences in expression that arise from alternative realizations of the arguments of verbs and other constituents (Katz & Levin, 1988). For example, suppose START is presented with a statement

```
Greece surprised the European Union with its actions.
```

This statement can also be paraphrased as "*Greece's actions* surprised the European Union." In order to match questions related to this alternate version of the statement, START must make use of a structural transformation rule that can be expressed as follows:

```
If        <<subject verb object1>with object2>
Then      <object2 verb object1>AND
          <object2 related-to subject>
Where verb belongs to the emotional-reaction class
```

With the addition of this rule, START can answer not only questions like

```
Did Greece surprise the European Union with its actions?
Did Greece surprise the European Union?
```

but also, it can answer questions like

```
Did Greece's actions surprise the European Union?
Which country's actions surprised the European Union?
```

Note that, within START, structural transformation rules are typically associated with classes of verbs, rather than individual verbs. The above rule applies for all verbs in the *emotional-reaction* class, which includes "surprise," "anger," "embarrass," and others. A range of other verb classes suited to use in structural transformation rules may be found in Levin (1993).

A second, significant aspect of START's design is the use of *natural language annotations* (Katz, 1997). Natural language annotations are natural language phrases and sentences associated with segments of information, describing their content. When START matches a user request to a natural language annotation, it can then access the associated segment of information as a response to the user.

For example, an HTML fragment containing information about clouds on Mars may be annotated with the following English sentences and phrases:

```
clouds on Mars
Martian clouds are composed of water and carbon dioxide.
...
```

START parses these annotations and stores the parsed structures (nested ternary expressions) with pointers back to the original information segment. To answer a question, the user query is compared against the annotations stored in the knowledge base. If a match is found between ternary expressions derived from annotations and those derived from the query, the corresponding annotated segment is returned to the user as the answer. For example, annotations like those above allow START to answer the following questions:

```
Are there clouds on Mars?
Do clouds exist on Mars?
What is the composition of Martian clouds?
Do you know what clouds on Mars are made of?
...
```

Figure 23.2 presents an example of START answering such a question.

Except for small amounts of particularly vital information, of course, it is impractical to annotate each item of content manually. However, sources of all types—structured, semistructured and unstructured—can contain significant amounts of parallel material. *Parameterized annotations* address this situation by combining fixed language elements with "parameters" that specify variable portions of the annotation. As such, they can be used to describe whole classes of content while preserving the indexing power of nonparameterized annotations. As an example, the parameterized annotation (with parameters in italics)

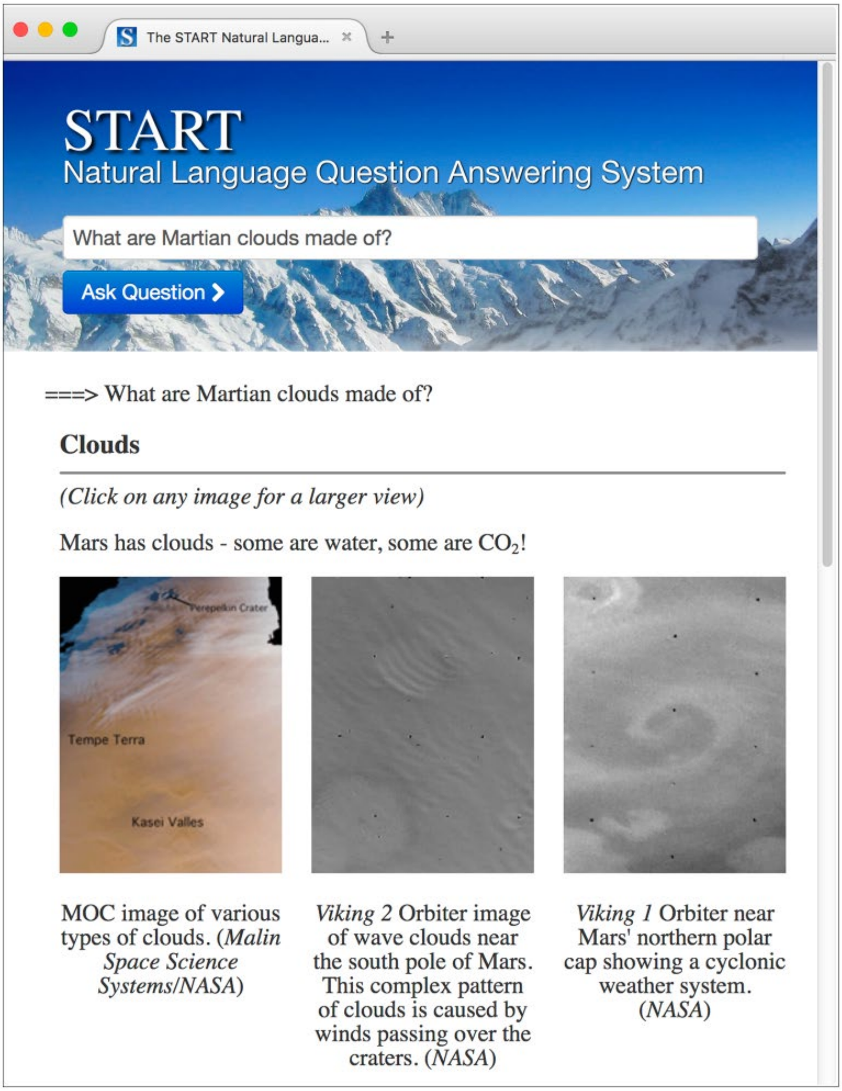


Figure 23.2 START answering a question using annotation-based matching.

```
number people live in city.
```

can describe, on the data side, a large semistructured Web resource containing population figures for various cities. On the question side, this annotation, supported by structural transformation rules, can recognize questions submitted in many forms:

```
How many people reside in Chicago?
Do many people live in Pittsburgh?
What number of people live in Seattle?
Are there many people living in Boston?
```

Additional parameterized annotations may be included that describe the population figures in other ways (for example, using the terms "population" or "populous"), and additional elements of the annotations may be parameterized. As a result, a large number of different questions can be answered using a small number of parameterized annotations. For example, with further parameterization, a single annotation can answer questions about area, elevation, population density, and other properties in addition to population.

The use of natural language annotations, and in particular, parameterized natural language annotations, enables START to respond to user requests in a wide variety of ways. For example, START can retrieve multimedia information or information from resources on the Internet, execute computations, retrieve foreign-language material, and perform specific actions on behalf of its user.

An important subset of retrievable information on the Internet and in structured datasets consists of data that may be viewed as collections of "objects," with each object having one or more "properties" that have particular "values." START operates in conjunction with a system called Omnibase that manages information that conforms to this *object–property–value* data model (Katz *et al.*, 2002).

Parameterized annotations serve as the interface between START and Omnibase's object–property–value data model, allowing the combined systems to answer questions about a variety of topics such as a country's population, area, GDP, or flag; a city's population, location, or subway map; or a famous individual's place of birth, date of birth, or spouse. The object–property–value data model is more generally applicable than it may appear on the surface, as many object–property questions can be cast with diverse phrasing; e.g., "What is Angela Merkel's date of birth?" can be phrased as "When was Angela Merkel born?" "What is Argentina's size?" can be phrased as "How big is Argentina?" and so forth. However, there are other possible types of queries that do not fall into the object–property–value model, such as questions about quantities of information that are a function of two objects (e.g., "How can I get from Boston to New York?"). Such questions and the information they request can be modeled through more general natural language annotations.

Figure 23.3 illustrates a question answered by START, utilizing support from Omnibase.

In order to match input questions to parameterized annotations successfully, START must know which terms can be associated with any given parameter. Omnibase supports this requirement by acting as an external gazetteer for source-specific terminology, with variants of terms being calculated from objects' names, extracted from semistructured material in information sources, or manually defined. This maintains the integrity of the abstraction layer: information source terminology is kept together with information source processing. Omnibase's use of the object–property–value data model applies equally well to fixed, semistructured websites and to "deep Web" sources that are accessed through special query languages or interactive form-based interfaces. When START transmits an object–property query to Omnibase, Omnibase executes an access script associated with the information source in question, and the access script may obtain individual elements of information directly from a static Web page, extract data from a local data source, or obtain dynamic information or otherwise "hidden" information by interacting with a query interface.

More recent work has made it possible for the START and Omnibase systems to access information without the need for manually created annotations. Many

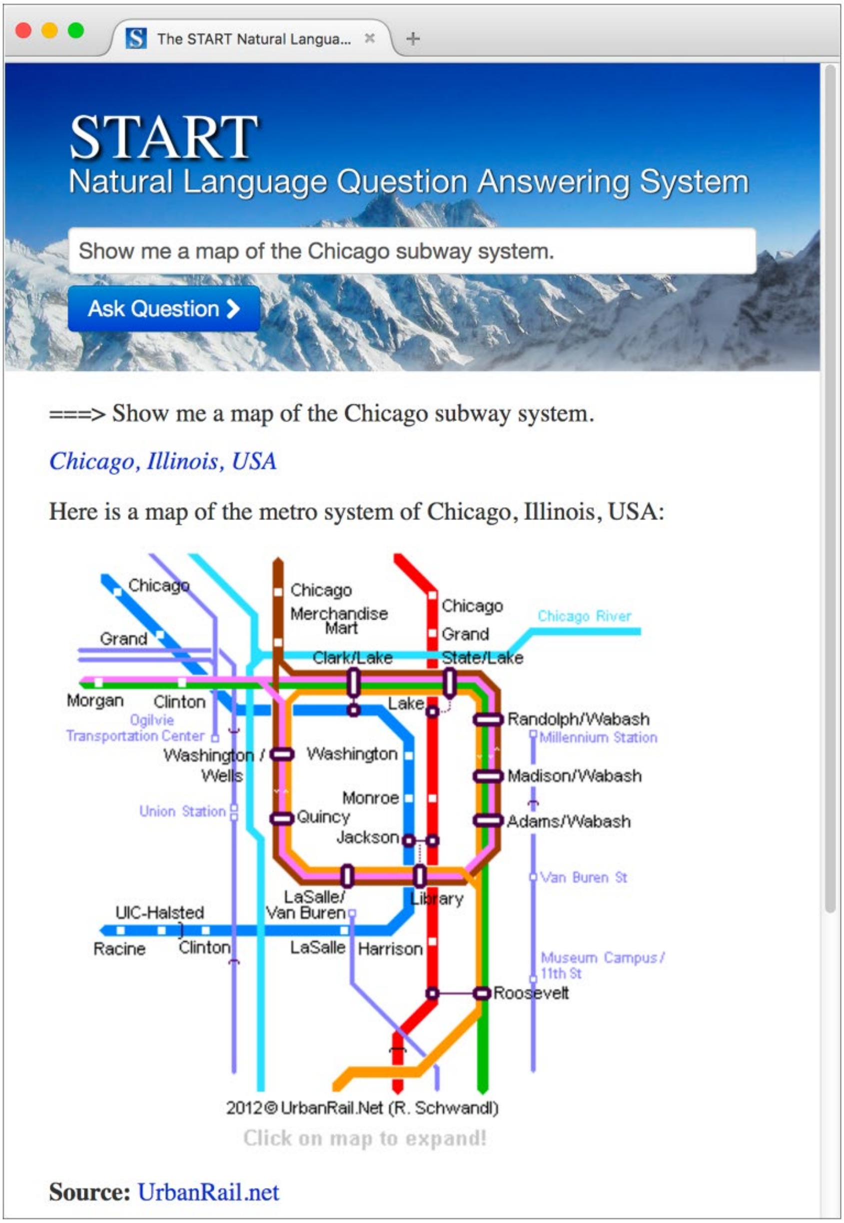


Figure 23.3 START and Omnibase answering a question using material from an English source.

information sources have a largely regular structure that enables us to extract property–value pairs. In addition, these property names are often in the form of English nouns and other phrases. In such cases, when an object–property–value question is asked, START first analyzes the question to find the object and property names, and then runs a procedure to extract the value as a response to the question. Using this technique, START can automatically answer questions such as "What was Einstein's alma mater?" or "What is the calling code of Italy?"

One particularly useful source of information is Wikipedia, the world's largest crowdsourced encyclopedia with over five million articles. Articles are organized in hierarchical sections, and many have an "infobox," a table that summarizes key information in the article. To access these kinds of information, we developed WikipediaBase (Morales, 2016), a system that turns Wikipedia into a virtual database and organizes it in an object–property–value data model. We consider infobox attributes and section headers to be properties. Using WikipediaBase, START is able to respond to requests like "Tell me about Albert Einstein's personal life" with the contents of the "Personal life" section, or "What awards did Einstein receive?" with the "Notable awards" row in Albert Einstein's infobox.

With its understanding of English morphology and syntax, START can recognize variations of object–property–value questions. It can correctly answer questions such as "Who designed the Oakland Bay Bridge?" (from the "Designer" property in the infobox) or "What river does the Brooklyn Bridge cross?" (from "Crosses: East River" in the infobox).

It is also possible to ask about information in ways that share little surface similarity with the property names. For instance, "Which college did Albert Einstein attend?" and "Where did Einstein study?" are valid paraphrases of "What is the alma mater of Albert Einstein?" but they share few content words. To address these types of questions, we compiled a crowdsourced corpus of over 15,000 questions about Wikipedia infoboxes. We used these questions to train a machine learning model that selects the correct response from a set of candidate answers with high accuracy (Morales, 2016; Morales, Premtoon, Avery, Felshin, & Katz, 2016). Our ongoing work in automatic techniques to answer questions will allow the START system to quickly scale up to new types of questions and information sources.

Some user requests may contain subrequests. For example, a user of START might submit a request "When was the president of France born?" Such questions are interesting because answering them typically involves information from different sources, and indeed, a system must answer one part of the question—e.g., "Who is the president of France?"—before proceeding to use the answer to that subquestion—in this case, François Hollande—within another subquestion to be answered—e.g., "When was François Hollande born?" Natural language annotations can help, in that they can be used to describe sets of simple questions that can be answered independently. In addition, the mechanism of parameter matching—via synonyms, hyponyms, etc.—plus the underlying mechanisms that supply answers to the simple questions, can be used to bridge terminology differences between resources, permitting a range of complex questions to be answered.

START utilizes an approach in which it analyzes complex questions linguistically in order to isolate valid candidate subquestions and determine an appropriate order in which to answer those subquestions. START then checks, via its base of annotated resource materials, to see if particular subquestions can be answered. This approach is described more fully in (Katz, Borchardt, & Felshin, 2005). Figure 23.4 provides an example of START answering a complex question.

START also contains a sophisticated natural-language-generation capability, which takes a set of ternary expressions as input and converts this set into readable English. In addition, prior to generation, ternary expressions can be joined together, modified, and augmented by the system, enabling START to produce individual sentences, narrative text, and dialog elements as appropriate.

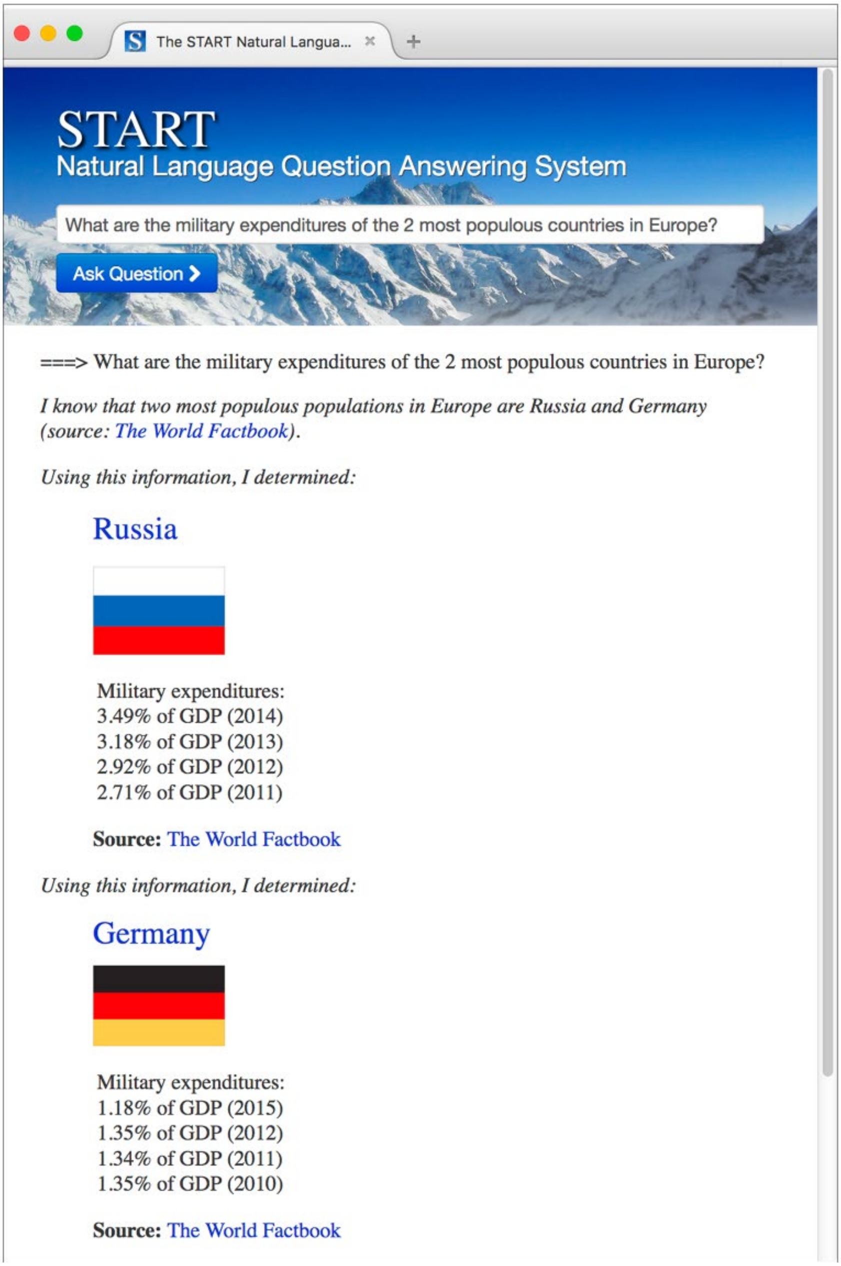


Figure 23.4 START answering a complex question using its syntactic decomposition strategy.

Taken together, START's collection of representations and techniques—its ternary expressions, structural transformation rules, natural language annotations, syntactic decomposition strategy, natural-language-generation capability, and so forth—provide a platform for interpreting a range of requests and issuing a range of responses to requests. Natural-language annotations, in particular, enable START to execute arbitrary procedures in response to requests, and this allows START to serve not only as a question answering system but also as an interface through which user requests can result in virtual or physical actions. The StartMobile system, described next, is one such application in which START is used to perform actions on a mobile device on behalf of its user.

StartMobile

In an application that foreshadowed the introduction of systems such as Apple's Siri, we used START to create a system called StartMobile, which provides a natural language interface to mobile devices (Bourzac, 2006; Katz, Borchardt, Felshin, & Mora, 2007; Katz, Mora, Borchardt, & Felshin, 2011). StartMobile allows its users to pose English requests for information present on their mobile devices, issue commands to perform actions on their devices, and make requests for information available from a broad range of sources beyond the confines of their device. Requests may be entered in written form, or by voice, using speech recognition utilities offered by Google, Inc.

StartMobile uses the START system as a first stage in the processing of user requests. START performs an initial interpretation of the requests, and if these requests concern the retrieval of general information from the World Wide Web or other sources, START obtains the information for presentation to the user. If it is not possible to complete the interpretation of the requests, however, or if the requests involve actions that must be performed on the user's mobile device, START encodes the user's requests in a language called Moebius, which has been designed to convey natural language requests in various stages of interpretation between systems and devices. Finally, software that resides on the user's mobile device completes the interpretation of user requests, if necessary, and performs required actions to fulfill those requests.

The StartMobile system supports a range of activities on several models of mobile phones:

- retrieving general-purpose information for the benefit of the user;
- retrieving contact and calendar information stored on the user's mobile device;
- retrieving text messages and managing the user's text message inbox;
- placing phone calls from the user's mobile device;
- creating reminders on the user's mobile device or on other users' mobile devices;
- taking pictures with the mobile device's camera;
- modifying device settings;
- accessing position information and displaying associated map and direction information on the user's mobile device; and
- retrieving video tutorials for presentation to the mobile user.

Figure 23.5 illustrates the StartMobile system in action, using typewritten entry of requests. (To place StartMobile in its appropriate historical context as a precursor to today's commercial systems, this and following screenshots show phone output captured in 2006–2007 during StartMobile's initial development.) In the interaction depicted in Figure 23.5, the mobile user has asked the system to list the user's contacts at a specific company.

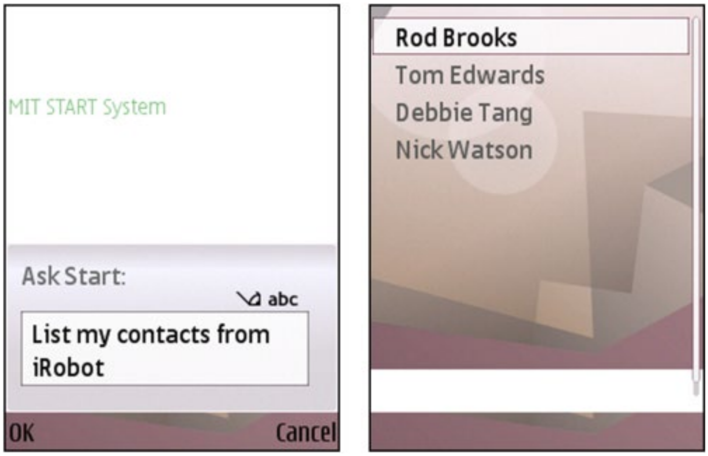


Figure 23.5 StartMobile performing a search within the contacts stored on a mobile device.

Through the use of parameterized annotations, structural transformation rules, and related technology, START enables the StartMobile application to accept requests in a range of variant forms. For the example illustrated in Figure 23.5, some of these variant forms are:

```
Who works for iRobot?
Who do I know at iRobot?
Which of my friends work at iRobot?
Show my colleagues from iRobot.
```

START is used for general information access in the StartMobile system. START's answers to general questions are streamlined for display on small screens, then relayed to the user's mobile device for presentation to the user. Figure 23.6 illustrates StartMobile used to retrieve multimedia information. An alternative mechanism within StartMobile allows users to submit text messages to the START server, then also receive responses by text message.

Figures 23.5 and 23.6 depict fully grammatical requests submitted by the user. StartMobile is also configured to allow the user to enter fragmentary utterances in a range of cases where the meanings remain clear. For the request illustrated in Figure 23.6, for example, the user could have entered "Helsinki subway map" or "subway map Helsinki," resulting in a display of the same map.

Other types of requests concern information maintained on the user's mobile device. To handle these requests, START matches them to natural language annotations as always; however, the annotated material in this instance is a procedure that relays instructions to the mobile device. Associated software on the mobile device performs the necessary operations and delivers the results to the user. Figure 23.7 depicts the handling of such a request, involving a search through the calendar on the user's mobile device.

Some user requests may contain unusual names—people, streets, cities, businesses, etc.—that appear within particular data entries on the user's mobile device. To enable START to correctly analyze these requests and take appropriate actions, StartMobile implements a mechanism whereby submitted user requests are initially inspected, on

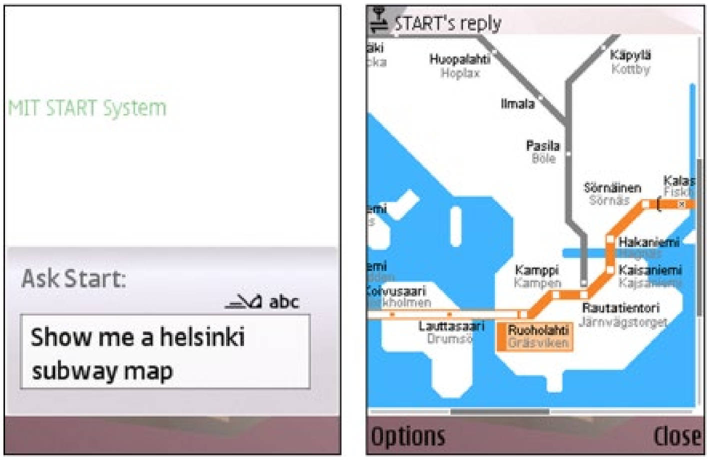


Figure 23.6 StartMobile responding to a request for general information.

Figure 23.7 StartMobile searching through the calendar on the user's mobile device.

the user's mobile device, to recognize and categorize names that appear in data sets such as the contacts database or calendar. START uses this information in a manner parallel to its use of the Omnibase system as a gazetteer. For the request illustrated in Figure 23.8, this mechanism enables StartMobile to correctly process the names "Boris" and "Federico."

In another set of cases, the user's input is not a request for information but rather a command to perform an action on the user's mobile device. These requests are handled in a similar manner to requests for information on the user's mobile device, with START relaying instructions to software that performs actions on the user's mobile device. Figure 23.9 illustrates StartMobile's handling of a request to take a picture using the camera on the user's mobile device.

In still other cases, the user may enter a request on one mobile device to perform an action on another mobile device. In this instance, START will relay instructions to

Figure 23.8 StartMobile presenting directions from one location to another location.

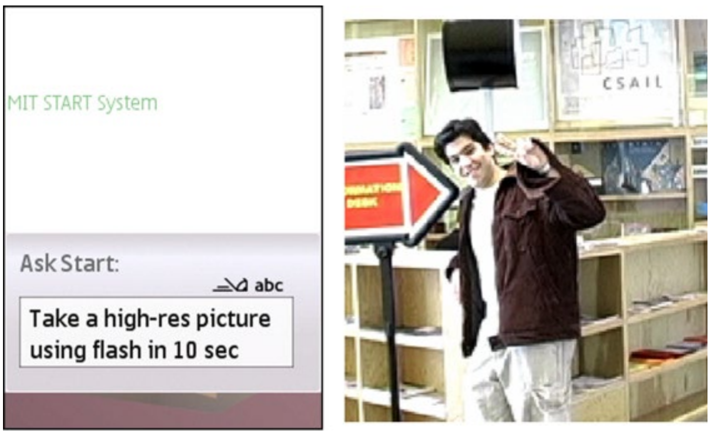


Figure 23.9 StartMobile responding to a request to take a picture using a phone's camera.

the first device, which must then relay appropriate instructions to the affiliated device. Figure 23.10 presents an example of this type of request being handled by StartMobile.

Within the StartMobile application, high precision is extremely important. The system is frequently asked to perform actions that may not easily be rescinded. Separately, in some circumstances interactivity may be less desirable than for, say, a user interacting through a computer console. Finally, limited display space increases the inconvenience caused by inappropriate responses in listed results.

Another significant issue for natural-language interfaces to mobile devices arises from the distributed nature of processing in this context. It is often the case that natural language requests can only be fully understood—their ambiguities resolved—in the presence of specific, matching components of knowledge. In distributed environments, this knowledge is distributed, requiring the networked devices and systems in some cases to collaborate not only toward the ultimate satisfaction of the received requests but also toward the initial understanding of requests, so that it is possible to satisfy them.

StartMobile makes use of an intermediate, language-based representation called Moebius (Borchardt, 2014), which supports distributed interpretation and distributed fulfillment of natural-language requests. Moebius serves to encode natural-language requests at varied stages of syntactic and semantic interpretation, so that these requests may be relayed between systems—for instance, a user's mobile device, central

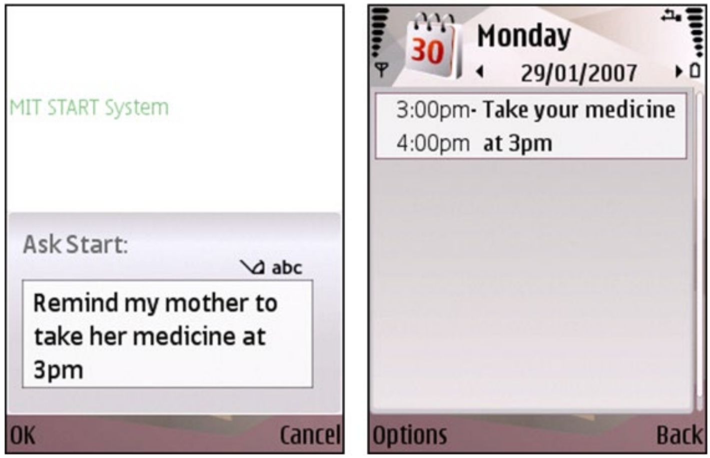


Figure 23.10 StartMobile posting a reminder on an affiliated mobile device.

servers, and other users' mobile devices—in order to receive additional interpretation and fulfillment. While Moebius specifically addresses the representation and processing of ambiguous requests, it is also applicable for more straightforward requests and thus we use the language as an intermediate representation for all StartMobile requests that must be relayed between systems and devices.

Following is an example of a Moebius expression issued by START to the user's mobile device, depicting a substantially interpreted version of the English request illustrated in Figure 23.10.

```
alert(object:person mother(of:person "user"),
      with:message_string
        "Take your medicine at 3pm.",
      at:time "2007-01-29T15:00:00")!
```

Moebius specifies basic syntactic relations between elements of the representation, and it adds semantic labels, drawn from a hierarchy of general to specific categories.

A key aspect of Moebius is that it uses language itself as a representation. In this respect, it shares a common orientation with the START system. START uses language-based ternary expressions as a representation for both questions and natural language annotations. Indeed, when START matches a question to a natural language annotation it does two things: it provides an answer to the question, and it commits to an interpretation of the question. Moebius can be thought of as extending this idea to distributed contexts, enabling partially interpreted requests to be interpreted and fulfilled by collective action on the part of multiple systems in a distributed environment.

As an example of the use of Moebius to characterize an ambiguous request at different stages of interpretation, consider the request

```
Is Carl at IBM?
```

This question could be offered to ascertain whether or not Carl is employed by IBM, or it could be offered to determine whether or not Carl is, at the moment, physically present at an IBM facility. We assume that the human user has constructed the request in compliance with conversational maxims such as proposed by Grice (1975)—that is, by supplying an adequate amount of information, but not too much information, by being truthful, by supplying only relevant information, and by being clear or perspicuous. To the human user, the request "Is Carl at IBM?" may be unambiguous in context; however, the system may need additional knowledge to disambiguate the request. The system may obtain this knowledge by consulting the repertoire of capabilities offered by components expected to fulfill the request (that is, whether these components are known to be able to respond to one interpretation or the other), by referencing contextual information from the current state of processing, by consultation with the human user, and so forth.

If the device that initially processes the request "Is Carl at IBM?" does not have access to the knowledge needed to fully interpret the request, then, using Moebius, that device can encode the request in a partially interpreted form:

```
be(subject:person "Carl", at:object "IBM")?
```

This representation parses the request syntactically, yet it makes no commitment as to the semantic interpretation of the relationship between "Carl" and "IBM" or as to the specific semantic category of "IBM" ("object" being the most general semantic category). If this request is relayed to another device or system that possesses the necessary knowledge to disambiguate the request, that system may cast the request into one of two more fully interpreted forms. If the determination is made that the request concerns physical presence at an IBM facility, the request can be reexpressed as

```
be(subject:person "Carl", at:facility "IBM")?
```

where "IBM" is classified semantically as a physical "facility." On the other hand, if the determination is made that the request concerns employment, the request can be reexpressed as

```
employ(subject:organization "IBM",
       object:person "Carl")?
```

where "IBM" is classified semantically as an abstract "organization," and the relationship is reexpressed as one of employment. Subsequent processing can then continue according to the chosen interpretation.

In StartMobile, when the mobile device receives the partially interpreted form of the request "Is Carl at IBM?" it chooses to interpret this as a request about employment. As a result, it translates the received expression into the more fully interpreted Moebius expression requesting employment information for Carl, then processes this Moebius expression. Figure 23.11 illustrates StartMobile's handling of the request "Is Carl at IBM?" using employment information recorded in the contacts database of the user's mobile device. The displayed output in this case explicitly informs the user that StartMobile has retrieved employment information, so as to clarify StartMobile's interpretation of the user's request.

In general, ambiguity can arise from many sources—abstract verbs; syntactic ambiguities; omitted adverbial phrases; ambiguous prepositions and conjunctions; abstract semantic categories; descriptions of objects; ambiguous names, times, and places; anaphora; and abstract adjectives and adverbs, to name a few. Moebius provides several

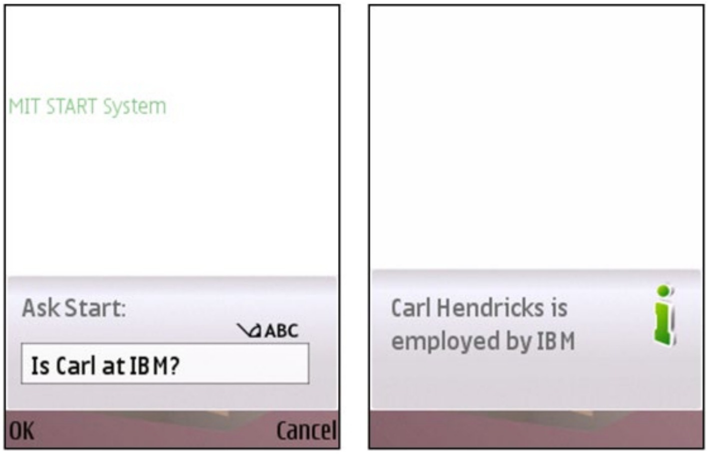


Figure 23.11 StartMobile responding to an ambiguous request.

mechanisms for depicting and resolving these ambiguities. Abstract verbs can be replaced by more specific verbs. For example, a request to "contact" a person can be reexpressed as a request to "call" a number or "send" a message. Abstract semantic categories can be replaced by more specific categories. For example, "message" can be replaced by "e-mail_message" "text_message" "voice_message" and so forth. Descriptive subexpressions such as "address(of:facility apartment(of:person "Sandra"))" and "3 o'clock" can be replaced by more specific expressions such as "298 Beacon Street, Boston, MA 02116" and "2013-07-22T15:00:00" In addition, ambiguous commands, statements, and questions can be clarified by inserting adverbial phrases, for example, or replacing the original expressions with entirely different expressions. The goal of Moebius is to capture a range of such ambiguities in various stages of interpretation, and we have found that simple natural language, structured for ease of computer processing, provides sufficient expressiveness to model many common ambiguities.

The StartMobile system situates START as a central server, accessed by one or more mobile devices that occasionally interact directly with one another. START performs initial processing of user requests, then passes Moebius requests to other systems and devices as needed for further interpretation and / or fulfillment. However, the overall design of the StartMobile system allows for other configurations as well.

One alternative is to perform initial processing of natural language requests on each user's mobile device. Each mobile device could then relay Moebius requests to other devices and systems as necessary for further interpretation based on knowledge held by those systems, or for the completion of actions external to that mobile device.

A third possibility involves a mixture of these two approaches. In this configuration, each user's mobile device would contain a lightweight capability for simple language processing, then pass off partially interpreted requests or even uninterpreted requests to other, more substantial language processing components as the need arises.

A particular emphasis of StartMobile has been enabling users to pose requests that result in actions: setting reminders, for example, or taking pictures or setting location-triggered alerts. Subsequent to the StartMobile effort, we have continued our exploration into the construction of natural language interfaces that carry out actions on behalf of their users. In particular, we used START as one component in a system called the Analyst's Assistant, which supports collaborative user–system interpretation of vehicle events exhibited in a dataset of vehicle track information (Borchardt *et al.*, 2014). From this and related efforts, we have arrived at the view that natural language interfaces that perform actions on behalf of their users can benefit greatly from targeted support for performing *collections* of interrelated actions. In the mobile phone domain, for example, a user might wish to use the interface to take a picture, then send it to a friend, then attach a label to the picture and save it in the phone's memory. Particular capabilities that can enhance the effectiveness of an interface in supporting these kinds of interactions are: (a) having a robust referencing capability, where the user can refer to previously mentioned quantities or previous responses of the system using simple constructions like "that picture," "my church friends," or mouse / touchscreen selections, and (b) providing support, both in input and output, for requests and descriptions at multiple levels of granularity, from coarse-granularity actions and returned summaries to fine-granularity specifications and "drill-down" results.

Commercial Systems

START and StartMobile are largely rule-based in construction, with these rules created through some amount of human involvement and effort. This enables the systems to respond to rather complex requests, well beyond the range of entity description and relationship requests that can be supported by simple keyword-based or statistical interpretation techniques. This has also enabled these systems to achieve very high precision in their responses. On the other hand, with the explosion of information available on the Web and maintained within mobile devices, it is difficult to provide comprehensive coverage of available sources and request types without some amount of automated construction of capabilities. Current commercial systems such as Apple's Siri, IBM's Watson, Google's "Google Now," Microsoft's Cortana, and Amazon's Alexa employ technology of the sort contained in START and StartMobile in combination with statistical interpretation and calculation of responses based on large-scale machine learning. This provides additional coverage for these systems, plus increased ability to handle ill-formed or idiosyncratic requests, at the expense of some amount of precision in the production of responses.

Current commercial systems may accept either written requests, speech input, or both. Systems that accept both written and speech inputs are typically designed in a modular fashion, as is the case with StartMobile, where speech input is independently processed and the results of speech recognition are submitted to the question answering component of the system.

Capabilities offered by Google, Inc., can serve as an illustration of current design practice in the construction of information access and question answering systems. While these capabilities provide broad coverage by using statistical machine learning techniques to match requests directly to potentially relevant material in unstructured sources, it is also the case that considerable attention has been given to *a priori* structuring of knowledge found in various sources, similar in spirit to the object–property–value structuring of knowledge in START's companion system Omnibase, leading to higher precision answering of particular types of requests (Dong *et al.*, 2014).

Conclusion

Cellular telephones and other mobile devices have the potential to provide users with many useful features and capabilities, but the more capable these devices become, the harder it becomes to make use of them with traditional interfaces. Natural language can express a wide range of requests in a very compact form, intuitively usable by humans. Natural language interfaces thus have the potential to significantly reduce the complexity of interaction with mobile devices. We view recent advances in the construction of natural language interfaces as welcome steps towards this goal.

A number of challenges remain in the design and construction of natural language interfaces to mobile devices:

- Further integration is needed between rule-based processing techniques, such as those that produce the high precision of responses in START and StartMobile, with large-scale machine learning technology, such as that which produces the domain coverage of current commercial systems.
- Many systems exhibit limited coverage of requests involving complex syntactic constructions, assumed context, and special-purpose vocabulary.
- Where speech input is accepted, tighter integration of the speech recognition and request processing components may be possible, so that the interpretation of speech input can to a larger extent be influenced by capabilities and constraints of the request processing component.
- Current systems rarely exhibit the ability to explain the manner in which their responses have been determined. Such explanations would help users assess the likelihood that a system's responses are appropriate to the submitted request.
- Finally, additional work is required in supporting distributed processing of requests, where multiple devices and systems hold pieces of information that are needed for interpretation and fulfillment of the requests.

Acknowledgements

The work described in this chapter has been supported in part through funding provided by the Defense Advanced Research Projects Agency, Nokia Corporation, and the Intelligence Advanced Research Projects Activity; in part by AFRL contract No. FA8750-15-C-0010; and in part by the Center for Brains, Minds, and Machines (CBMM), funded by NSF STC award CCF-1231216. The authors also wish to thank Alvaro Morales for assistance with this chapter.

References

Borchardt, G. C. (2014). *Moebius language reference, version 1.2* (Report MIT-CSAIL-TR-2014-005). Cambridge, MA: MIT Computer Science and Artificial Intelligence Laboratory.

Borchardt, G., Katz, B., Nguyen, H.-L., Felshin, S., Senne, K., & Wang, A. (2014). *An analyst's assistant for the interpretation of vehicle track data* (Report MIT-CSAIL-TR-2014-022). Cambridge, MA: MIT Computer Science and Artificial Intelligence Laboratory.

Bourzac, K. (2006, April 27). Nokia phones go to natural language class. *Communications News, MIT Technology Review*. Retrieved from http://www.technologyreview.com/news/405713/nokia-phones-go-to-natural-language-class/

Dong, X., Gabrilovich, E., Heitz, G., Horn, W., Lao, N., Murphy, K.,… Zhang, W. (2014). Knowledge Vault: A web-scale approach to probabilistic knowledge fusion. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 601–610). New York, NY: ACM.

Grice, H. P. (1975). Logic and conversation. In P. Cole and J. L. Morgan (Eds.), *Syntax and Semantics, Volume 3: Speech Acts.* New York, NY: Academic Press.

Harabagiu, S. M., Maiorano, S. J., & Pasca, M. A. (2003). Open-domain textual question answering techniques. *Natural Language Engineering*, *9(3)*, 1–38.

Hardesty, L (2011, April 19). The brains behind Watson. *MIT News Magazine, MIT Technology Review.* Retrieved from http://www.technologyreview.com/article/423712/the-brains-behind-watson/

Hendrix, G., Sacerdoti, E., Sagalowicz, D., & Slocum, J. (1978). Developing a natural language interface to complex data. *ACM Transactions on Database Systems*, *3(2),* 105–147.

Katz, B. (1980). *A three-step procedure for language generation* (A.I. Memo 599). Cambridge, MA: MIT Artificial Intelligence Laboratory.

Katz, B. (1988). Using English for indexing and retrieving. In *Proceedings of the 1st RIAO Conference on User-Oriented Content-Based Text and Image Handling (RIAO '88)* (pp. 313–333). Paris: Centre de Hautes Etudes Internationales d'Informatique Documentaire (CID).

Katz, B. (1990). Using English for indexing and retrieving. In P. H. Winston & S. A. Shellard (Eds.), *Artificial intelligence at MIT: Expanding frontiers* (vol. 1, pp. 134–165). Cambridge, MA: MIT Press.

Katz, B. (1997). Annotating the World Wide Web using natural language. In *Proceedings of the 5th RIAO Conference on Computer Assisted Information Searching on the Internet (RIAO '97)* (pp. 136–155). Paris: Centre de Hautes Etudes Internationales d'Informatique Documentaire (CID).

Katz, B., Borchardt, G., & Felshin, S. (2005). Syntactic and semantic decomposition strategies for question answering from multiple resources. In *Proceedings of the AAAI 2005 Workshop on Inference for Textual Question Answering* (pp. 35–41). Menlo Park, CA: AAAI Press.

Katz, B., Borchardt, G., & Felshin, S. (2006). Natural language annotations for question answering. In *Proceedings of the 19th International FLAIRS Conference (FLAIRS 2006)* (pp. 303–306). Menlo Park, CA: AAAI Press.

Katz, B., Borchardt, G., Felshin, S., & Mora, F. (2007). Harnessing language in mobile environments. In *Proceedings of the First IEEE International Conference on Semantic Computing (ICSC 2007)* (pp. 421–428). Los Alamitos, CA: IEEE Computer Society Conference Publishing Services.

Katz, B., Felshin, S., Yuret, D., Ibrahim, A., Lin, J., Marton, G.,… Temelkuran, B. (2002). Omnibase: Uniform access to heterogeneous data for question answering. In *Proceedings of the 7th International Workshop on Applications of Natural Language to Information Systems (NLDB 2002)* (pp. 230–234). Berlin: Springer.

Katz, B., & Levin, B. (1988). Exploiting lexical regularities in designing natural language systems. In *Proceedings of the 12th International Conference on Computational Linguistics (COLING '88)* (pp. 316–323). Budapest: John von Neumann Society for Computing Sciences.

Katz, B., Mora, F., Borchardt, G., & Felshin, S. (2011). *StartMobile: Using language to connect people to mobile devices* [Video File]. Retrieved from https://www.youtube.com/watch?v=BqOKqXaUWOw.

Levin, B. (1993). *English verb classes and alternations: A preliminary investigation*. Chicago, IL: University of Chicago Press.

Morales. A. (2016). *Learning to answer questions from semi-structured knowledge sources* Master's Thesis, Massachusetts Institute of Technology, Cambridge, MA. Retrieved from https://dspace.mit.edu/handle/1721.1/105973

Morales, A., Premtoon, V., Avery, C., Felshin, S., & Katz, B. (2016). Learning to answer questions from Wikipedia infoboxes. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing (EMNLP 2016)* (pp. 1930–1935), Stroudsburg, PA: Association for Computational Linguistics.

Murdock, J. W. (2012). This is Watson. *IBM Journal of Research and Development*, *56*(3–4).

Nyberg, E., Burger, J., Mardis, S., & Ferrucci, D. (2004). Software architectures for advanced QA. In M. Maybury (Ed.), *New directions in question answering*. Cambridge, MA: MIT Press.

Waltz, D. L. (1978). An English language question answering system for a large relational database. *Communications of the ACM*, *21*(7), 526–539.

Weischedel, R., Xu, J., & Licuanan, A. (2004). A hybrid approach to answering biographical questions. In M. Maybury (Ed.), *New directions in question answering*. Cambridge, MA: MIT Press.

Woods, W. A., Kaplan, R. M., & Nash-Webber, B. L. (1972). *The Lunar Sciences Natural Language Information System: Final report* (Report No. 2378). Cambridge, MA: BBN Technologies.

24

Visual Query Interfaces

Tiziana Catarci, Massimo Mecella, Stephen Kimani, and Giuseppe Santucci

Introduction

A *visual query system* (VQS) can be defined as a system that uses a visual representation for both the domain of interest and requests related to the domain. The first graphical query language, which was referred to as Query-By-Example (QBE), was introduced in the mid-1970s (Zloof, 1997). A wide range of implementations were built using QBE concepts and there are several tools using this paradigm today.

Since the purpose of any VQS is to provide access to the information contained in a database, the main users' tasks are *understanding the database content*, focusing on meaningful items, *finding query patterns*, and *reasoning about the query result*. These tasks require specific techniques to be effectively accomplished, and such techniques involve activities such as pointing, browsing, filtering, and zooming—all activities that nicely fit with visual representations and direct manipulation mechanisms. For instance, if the result of an information request can be organized as a visual display, or a sequence of visual displays, the information throughput is immensely superior to the one that can be achieved using only textual support, and the users can directly point at the information they are looking for, without any need to be trained in the complex syntax of query languages. Alternatively, users can navigate in the information space, following visible paths that will lead them to the target items. Again, thanks to the visual support, users are able to understand easily how to formulate queries and they are likely to achieve the task more rapidly and less prone to errors than with traditional textual interaction modes.

In modern VQSs, information visualization (a.k.a. *infovis*) mechanisms are used for displaying the query results and making sense out of them. Indeed, infovis relies on basic features that the human perceptual system inherently assimilates very quickly: color, size, shape, proximity, and motion. These features can be used by the designers of information systems to increase the data density of the information displayed. Because the users perceive such features so readily, and because each feature can be used to represent different attributes of data, good visualizations enable them to not only perceive information more easily but also to perceive more information at one time. In this way,

The Wiley Handbook of Human Computer Interaction, Volume 2, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

visualizations can reduce the search for data by grouping or visually relating information. While visualizations compact information into a small space, they can also allow hierarchical search by using overviews to locate areas for more detailed search. In fact, they also allow zooming in or popping up details on demand. Infovis, through aggregation and abstraction, enables users to recognize gaps in the data immediately, discover outliers or errors in the data, pinpoint minimum and maximum values, identify clusters, compare objects, visually draw some conclusions, and discover trends and patterns.

In order for a VQS to help users perform the tasks they have in mind, it has to be usable. Usability is a major criterion in assessing the quality of interactions between the user and the overall system. Usability of VQSs was first studied through a comparison of QBE and SQL (Reisner, 1988, Ziegler & Fahnrich, 1988). Reisner (1988) showed better user performances when using QBE with respect to SQL, both in query reading and query writing tests. However, the study by Ziegler and Fahnrich (1988), also comparing QBE and SQL, took into account several factors, such as the use of the same database management system and whether there was a similar environment. It is interesting to note that the query language type affected user performance only in "paper-and-pencil" tests, in which case QBE users had higher scores than SQL users. In online tests, the user's accuracy was not affected by the type of the language adopted, but the user's satisfaction was much greater with QBE, and his or her efficiency much better.

More recently, SQL and QBE were again compared through an experiment (Hvoreckýa, Drlikb, & Munk, 2010). The authors found that the time required for query formulation applying a QBE-based approach was shorter than the time required using an SQL approach; the participants were also more comfortable during the creation process while adopting the QBE paradigm instead of the traditional SQL approach. Interestingly, there were no remarkable differences regarding the accuracy of the queries between the two approaches. Even considering some limitation of the experiment (e.g., the choice of the participants can be questionable, QBE does not cover all the possible visual approaches and visual query languages properties), it is again quite clear there are some advantages for using a visual approach while generating queries.

In this chapter, we will first review the main approaches to visual queries, and later consider recent advancements specifically in the field of data-stream processing.

Classifying Visual Query Systems

A general overview and classification of VQSs can be found in Catarci, Costabile, Levialdi, and Batini (1997). According to the visual representation adopted for the database and the queries, VQSs are categorized into form based, diagram based, icon based, and a combination of these.

A form is a generalization of a table, and it is possible to represent relationships among cells, a subset or the overall set, allowing a three-level answer. There are VQSs in which it is possible to manipulate both the intensional and extensional parts of the database, focusing on different parts of the database.

Diagrams are frequently used in VQSs, which generally use some visual components (e.g., shapes, colors, arrows) that are univocally mapped into a concept.

In an icon-based system, there is a mapping between a real concept or analogy and an icon that hides the schema of the data. It is possible to query the database by combining icons according to spatial concepts. The main problem in designing an iconic system is to define an unambiguous mapping. While different attempts are made to find a common mapping, still there are no universal standards.

Another possible categorization is made by considering the strategies to understand the reality of interest. The filtering of the information of interest can be accomplished using a top-down strategy. The implementation can be made in several ways: iterative refinement, selective hierarchical zoom, or user-system dialogue. A different approach is browsing, which enables getting more knowledge by exploring the neighborhood concepts. Browsing can take various forms including extensional browsing, intensional browsing, or mixed browsing. An alternative approach is schema simplification, which "brings the schema close to the query." This can be realized through transformations of concepts of the original schema in a user view, which cannot be extracted by the original schema. Transformations are made to produce better query representation.

The visual query languages (VQLs) are also classified according to the query formulation strategy. In a schema navigation strategy, the user starts from a concept and can reach the other concepts of interest. There can be different paths to navigate the schema. The first possibility is to use an arbitrary path to explore the schema, reach the concepts of interest and apply condition(s) to them. It is also possible to select one concept from the database and then navigate the schema by a hierarchical view built using the chosen concept as a root. Moreover, users can choose the starting concept and then build their own relationships.

A second possible strategy in the query formulation process is by using subqueries. This can be accomplished using the following two approaches: by composition of concepts, usually in iconic-based languages in which several icons are combined to write the final query, or using stored queries previously created or stored in a system library.

Another strategy for query formulation is by matching, which can be done by example or by pattern. In a matching-by-example strategy, users can provide an example of a query and the system generalizes the example and builds the query. In a pattern-matching strategy, the system searches in the database for a pattern specified by the user.

The last strategy for query formulation is by using range selection. In this strategy, it is possible to specify a range on different dataset through graphical widgets.

In the following, we will present some prominent approaches. Table 24.1 reports a summary view of them, by classifying them according to the previously introduced dimensions.

Table 24.1 Classification of approaches.

Approach	*Visual representation*	*Strategy for understanding the reality*	*Strategy for query formulation*
QBD*	Diagram	Browsing	Subqueries
MURAL	Diagram	Browsing	Subqueries
QBB	Icon	Browsing	n/a
QBI	Icon	Filtering	Select project
Flow	Hybrid	Filtering	Design a flow
Kaleidoquery	Hybrid	Filtering	Design a flow
VISUAL	Hybrid	n/a	Hierarchical

Query-by-Diagram

*QBD** QBD* (Angelaccio, Catarci, & Santucci, 1990) is an example of a diagram-based VQL. This system is an entity-relationship (E-R) oriented data model, which provides a relationally complete query language. The graphical interface, which is the same for both schema specification and query formulation, relies on a language that also supports recursive queries. The main architecture is composed of three main modules: the graphic interface, the translator and the DBMS Interface. Users can interact with the graphic interface in four different ways: (a) by using the "E-R schema library," users can access the schemata of the applications, whereas (b) in the "E-R schema user library" there are the user views of the schemata stored in the schema library. (c) For each schema in the schemata library, there is also a set of schemata at a higher level of abstraction stored in the "E-R top-down schema library." (d) Users can store graphical queries in the "user query library" and reuse them when needed.

In the translator module, there is a translation from graphical queries into relational algebra, or into suitable programs if the original query is a recursive query. Then, the DBMS interface translates the relational algebra into a query in the underlying database language.

In QBD*, the query formulation process has different interchangeable steps. First, a user can explore the conceptual schema by a top-down browsing mechanism. Second, using graphical primitives called "location primitives" it is possible to focus on the subschema of interest. This can be done by direct extraction, expressing a query on the schema or using the schemata stored in the library. Further, it is possible to manipulate the schema by graphically replacing primitives, thereby bringing it "close to the query." After this transformation, the schema can result in a different schema, which is nonisomorphic to the original schema. Subsequently, the query is completed by graphical operations such as navigation or selection upon the database schema. In QBD* it is only possible to define queries on the database schema, which means at the intensional level. Figure 24.1 (a) shows a schema in the QBD* interface, in which the schema of a database dealing with *tickets* `concerning` *trips* `related to` *travels* `involving` *students* is shown. Figure 24.1 (b) shows the definition of a condition in the QBD* interface, in which the value `BOSTON` is used for the field `TO_CITY`, by drawing a line connecting them.

MURAL A similar approach to QBE can be found in the visual query language MURAL (Reiss, 2002), which is intended for multiple data sources integrating different types of data. The main objects are entities and relationships. The entities are all different data sources, and the relationships represent ways of correlating the data sources. An entity can represent a tuple from a relational database, an object from an object database, or a C++ or a Java object. Each entity has several fields defined over a set of domains like primitive type of data (e.g., string, integers) or references to other entities. A relationship can be a simple way to relate entities in one set with those in another set or a more complex relationship.

MURAL introduces several concepts in order to facilitate the creation of complex queries: the notion of combination of both entities and relationships to express AND, OR conditions, restrictions and fields. It is also possible to save a query for a submodel, so that users can instantiate such submodels as needed for building new complex queries. Figure 24.2 shows grouping operations in MURAL, in particular three entities (two classes and a method) are grouped together, as one of them is calling the other class from a method.

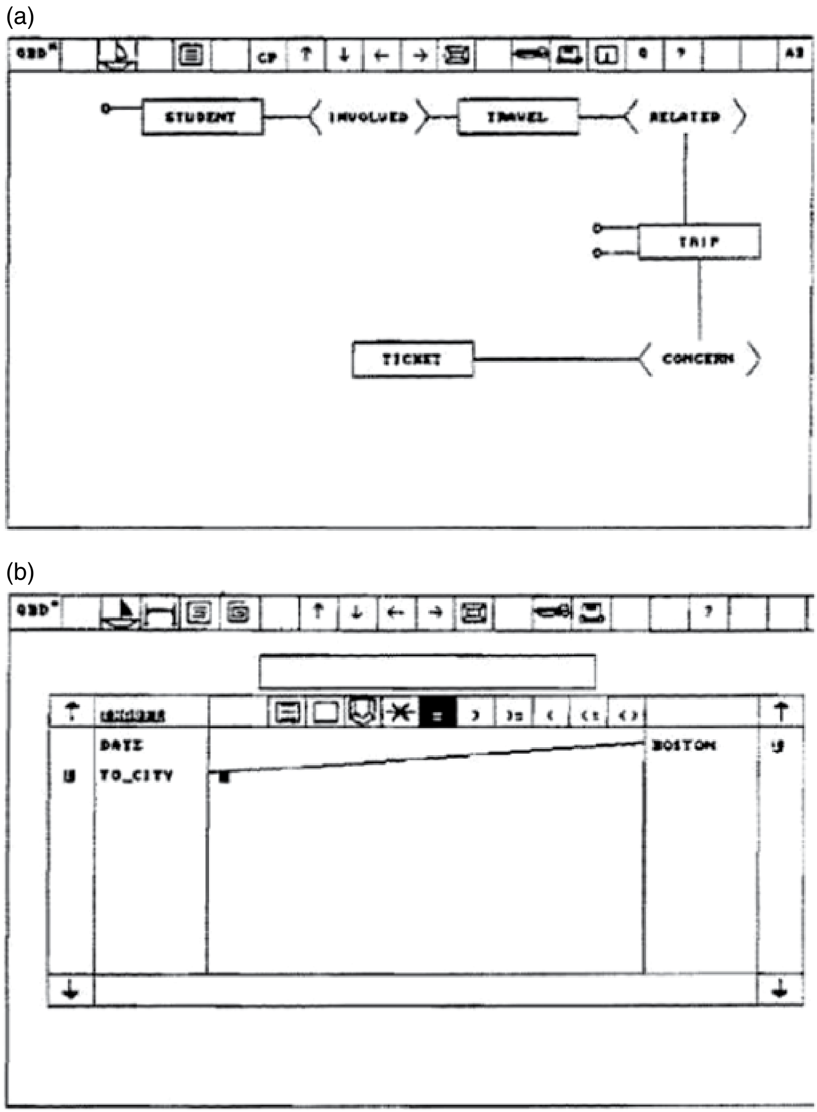


Figure 24.1 Screenshots of the QBD* interface. (a) Schema of interest. Rectangles represent the entities, and square brackets ("<" and ">") the relationships among entities. (b) Definition of a condition during query formulation.

Query-by-Icon (QBI)

QBB Query-by-Browsing (QBB) is a framework that allows both intensional and extensional queries (Polyviou, Evripidou, & Samaras, 2004). This framework adopts the same metaphor as most current operating systems, i.e., the desktop paradigm and related concepts like folders, documents, and applications. These objects are displayed in a tree mode, starting from a root folder with subfolders. In QBB, both the schema and query are represented by a folder hierarchy. The concept of a folder is strongly related to the

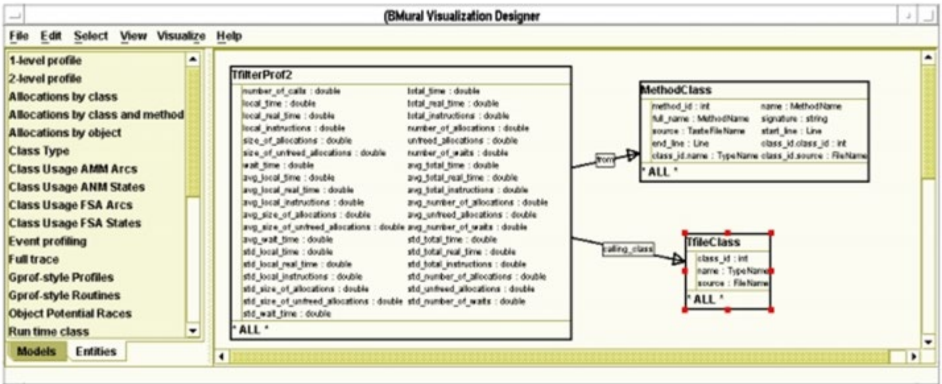


Figure 24.2 MURAL grouping operations: two classes and a method are shown, and their relationships (*calling*, *from*).

table of a database, and the subfolders are all the folders related to the parent folder. Documents and applications are views on the data. Documents are used to display the data whereas the manipulation of the data, such as insertion or deletion, is performed by the applications. Filtering, which is a special kind of an application, makes it possible to restrict the record in the parent folder. There are filter templates that are strongly related to the SQL predicates. However, it is also possible to build custom filters for other data types. In QBB the distinction between database navigation and query formulation is represented by the activation of a folder. Specifically, through implicit or explicit activation, it must be clear whether the folder is involved in a query or if it is only being browsed. Figure 24.3 shows the activation of a filter in QBB, in which Students are filtered on the basis of their `Level` being equal to `Postgraduate`.

QBI Query-by-Icon (Massari, Pavani, Saladini, & Chrysanthis, 1995) is a pure iconic VQL, which provides tools for the intensional browsing of databases. A user can formulate queries without having to know the underlying structure of the database and the path specification. A single icon hides the path expression that is automatically generated by QBI. The external view is made up of only two concepts, a class of objects and attributes of a class. The entire database is expressed by a set of classes with several properties called generalized attributes (GAs). A GA represents a generic property of one or more classes and can encapsulate both implicit and explicit relationships. Both classes and GAs are represented by icons. In order to avoid disambiguation, system-generated natural language descriptions are added to the icon visualization. The schema of the database is made according to a semantic data model called a graph model. The schema consists of a labeled graph that captures both structural information, such as classes and relationships, and consistency constraints. The classes of objects, which are nodes in the semantic model, are connected through paths. This concept is related to the GAs. However, not all the paths are equally meaningful. A QBI therefore defines a semantic distance function to estimate the meaningfulness of a path. This is done in order to only present the user with a restricted number of useful GAs, which would otherwise be shown in an infinite number. The query process formation follows the select-project paradigm. The user first defines the conditions that determine a subset of class and then

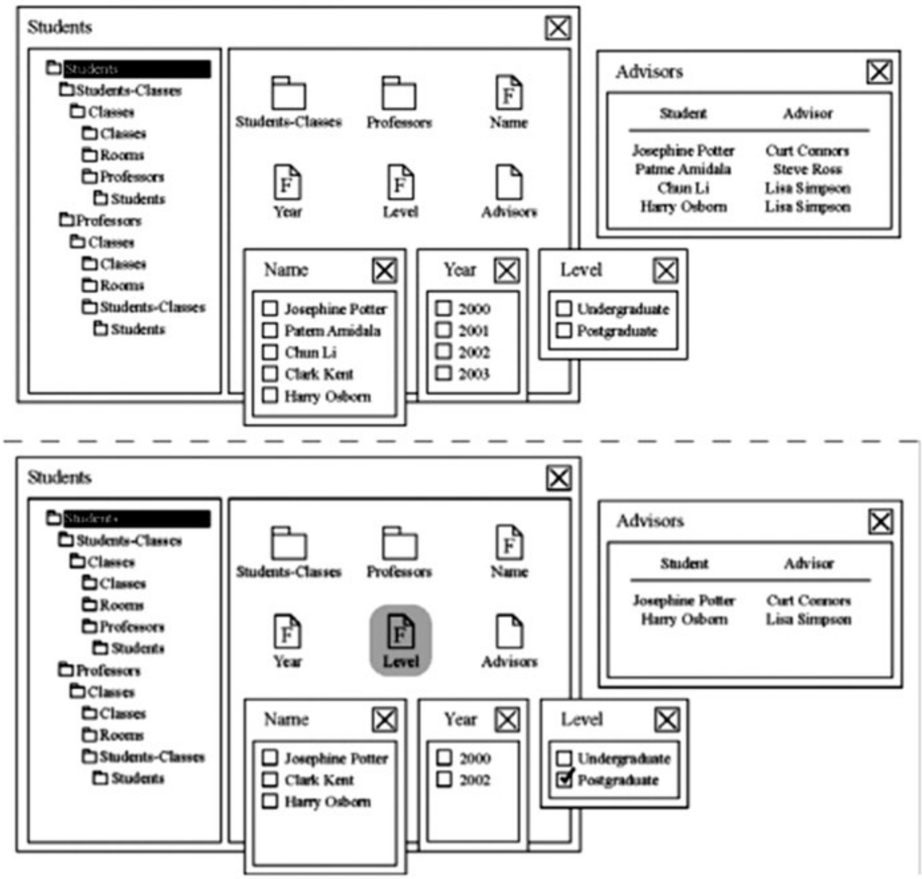


Figure 24.3 QBB—Activating a filter. Students have properties Name, Year, Level, and have relationship with Advisors. Possible values are shown in the popup windows.

specifies the GAs that will be part of the output result. Figure 24.4 shows the QBI interface, in which different icons are shown, in particular those representing the entities of a database dealing with university students. The reader may note the *Browser* window, showing some specific predefined queries (GAs), as well as the *Query Space* window, giving details on a specific GA in which a given condition is defined for the class professor.

The flow metaphor Morris, Abdelmoty, and El-Geresy (2002) designed a VQL for spatial databases. Such databases focus on representing and formulating queries about data related to objects located in the space. The language is valid for both spatial and nonspatial databases, and all the operations are expressed consistently. Some insights can be found in this implementation, where the most important is the metaphor used to define the query. Queries are visualized by a flow of information from the data source to the result. In between, there is a filter process where constraints can be applied. The flow starts with an icon representing an object. A simple filter expressing constraints can be applied and the flow will pass through the filters only as long as the constraints set are satisfied. Boolean conditions can be created by combining filters. The AND condition is represented by two filters in a series, whereas the OR condition

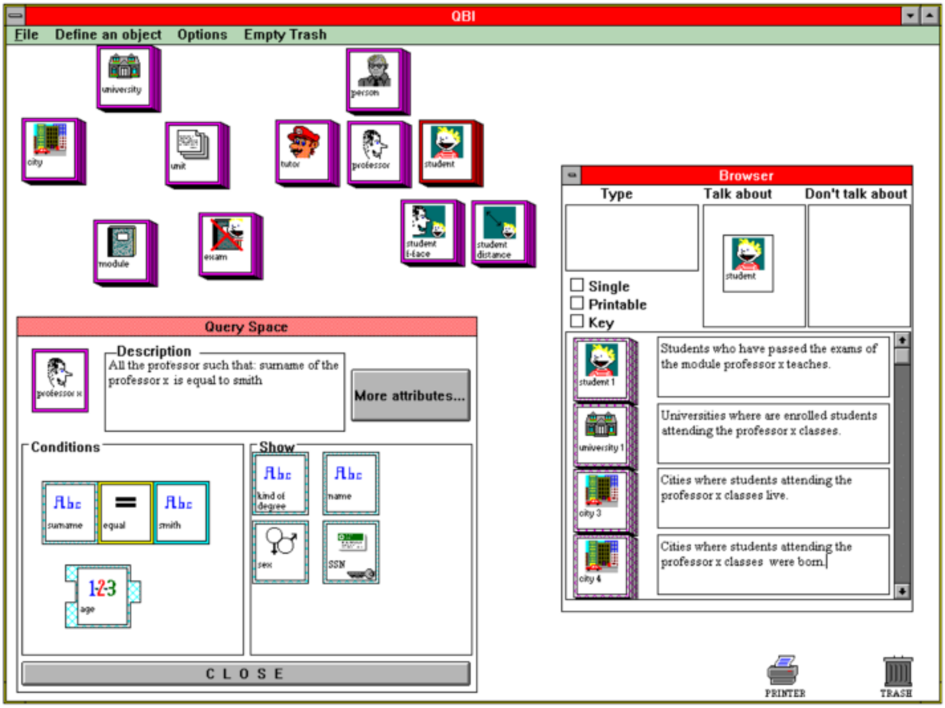


Figure 24.4 The QBI interface.

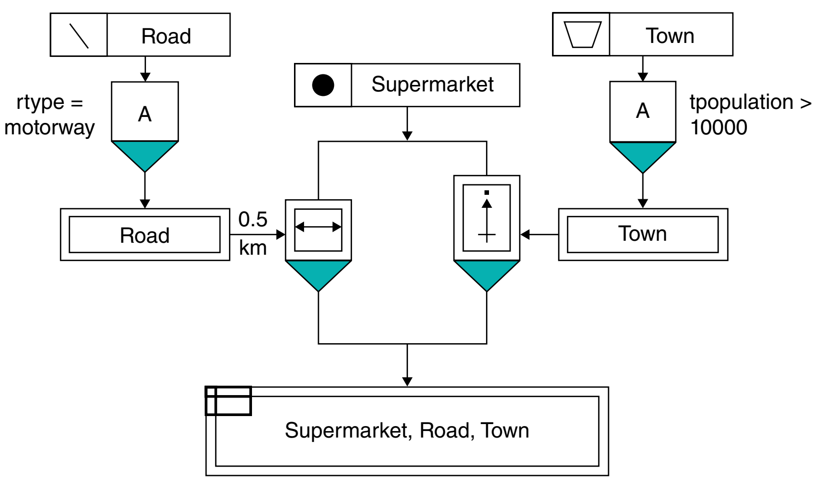


Figure 24.5 A spatial query.

is represented by two filters in parallel. If a filter has a double border it means a join condition, and this kind of icons is associated with more than one object type. Combining these basic constructs makes it possible to build complex queries. Figure 24.5 shows a spatial query based on the flow metaphor by Morris et al. (2002), attempting to select Supermarket around 0.5 km from a Road of type motorway or in a Town with a population greater than 10.000 (persons).

Kaleidoquery The flow metaphor is also the basic idea for the Kaleidoquery (Murray, Paton, & Goble, 1998), which is a visual query language for object databases. Like Morris et al.'s (2002) work, class instances and their extents enter the query, and where in the flowing process there are one or more filtering steps in which some constraints are applied on the attributes of the classes. The results of the query can be visualized or further used as input to start a new querying flow. Classes and extents are represented by a combination of icons and a textual description to give a better understanding than a pure iconic or pure textual visualization. As a user becomes more familiar with the system, the user will rapidly associate the icon with the text without fully reading it to understand the meaning of the query. The extents visualization consists of the extent name surrounded by an oval box. Kaleidoquery provides different icons to describe boolean operators, which can be easily combined using a parallel or serial connection as seen in Morris et al.'s (2002) work. It is possible to express basic constraints, such as: equal, greater than, and less than operators, adding them in the flow thereby restricting the result. Figure 24.6 show a complex query, in which the reader can easily identify the iconic visualization of the operators. Two extents involved in a join are identified with an equal condition applied on the attributes of the join. Aggregation operators, such as sum, maximum, average, are visualized by an oval surrounding the extents. The membership test and the universal and existential quantification are displayed with an oval arrow surrounding a textual description.

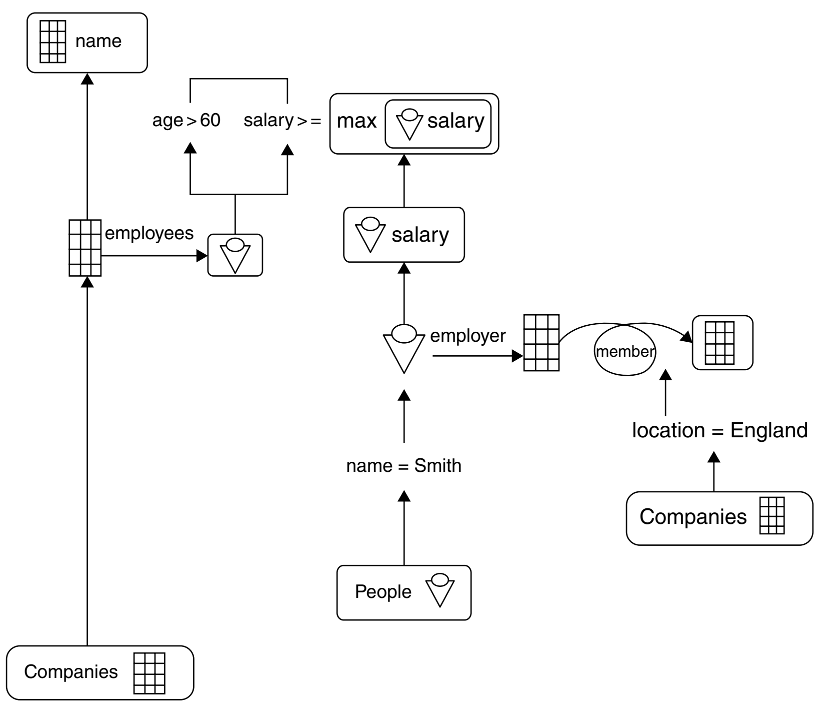


Figure 24.6 A graphical query in Kaleidoquery, in which, after selecting Companies located in England with employeer of name = Smith, and taking into account their salary (salaries), the user is asking for the name of employees aged more than 60 or having a salary greater than the max of Smiths' salary (salaries).

Kaleidoquery relies on OQL, an object query language used to work with query databases. One limitation of this language is that while writing the query, the user must also concentrate on the desired structure of the result. To facilitate the structuring process, the system allows the user to apply all the visualization conditions directly on the query results. This means that starting from the final extent of the flow, it is possible to apply grouping, order by, or other conditions, before visualizing the desired output.

VISUAL Another example of an icon-based object-oriented query language is VISUAL (Balkir, Ozsoyoglu, & Ozsoyoglu, 2002), which is a system addressing scientific databases. The system design is intended for handling large volumes of data with real-time constraints and spatial properties. The query part is implemented as an object. Processing the results can be communicated in between different query objects. Security, synchronization, and time-constraint issues are better managed in this object-oriented approach. In VISUAL there is a client-server approach for the query object model, where a query object acts as a client when requesting services from another query object, which becomes a server. Every query object is described using interpretation semantics. While there can be different execution semantics, objects can communicate through the interpretation semantics and can be executed in different frameworks. The query is represented by a window, as shown in Figure 24.7, which is divided into query head and query body. In the head there is a name of the query, input, and output parameters, and an output specification. VISUAL is strongly typed, and each output parameter must be specified as a single object or a collection of objects. The query body contains several iconized objects, condition boxes and links to other queries. Every iconized object has some properties, such as color or shape, which clearly identify it. There can be four classes of iconized objects: domain objects, method object, range object and spatial enforcement region object. VISUAL focuses on spatial and hierarchical concepts, in which objects can intersect each other or can be contained in another object specifying the relationships among them. The object oriented architecture allows an easy change of the domain of interest. The domain is the lowest layer of the system architecture. Thus, it is possible to build a new application by only writing at the lowest layer of the architecture.

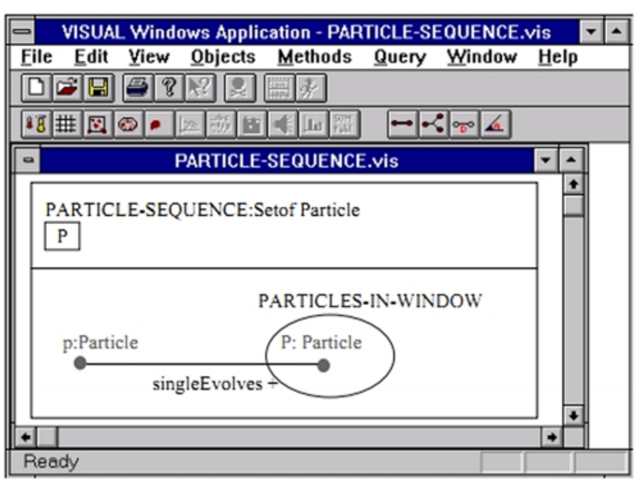


Figure 24.7 VISUAL representation.

Comparing VQLs

A VQL should provide different kinds of interaction because there is no unique paradigm that leads to the best results. An empirical experiment (Badre, Catarci, Massari, & Santucci, 1996) about the ease of use of two different query languages shows that there can be some advantages as well as disadvantages in both iconic and diagram-based approaches. In the experiment comparing QBI and QBD systems, different strategies are used for the query formulation (navigation versus composition) as well as some different visual formalisms (diagrams versus icons), which are basic aspects of a VQL. Thus, the results can be extended to larger classes of VQSs. The experiment focuses on discovering which relation occurs between the query language type and both the query class and the experience of the user. In particular, the queries were classified according to the semantic distance of the path involved in the query and the overall number of the cycles in the query, where the notion of path derives from the graph model described in QBI. The main result is that both accuracy and response time seem to be highly sensitive to the semantic distance of the query path, whilst QBD shows independence for both criteria. In addition, QBD is less accurate and requires more time when there are cycles in the query, and QBI seems not to be affected by the presence of cycles.

Recent Applications of VQSs

VQSs mainly deal with traditional databases, i.e., databases containing alphanumeric data. However, in recent years the application realms of databases have increased greatly in terms of both the number and variety of data types. As a consequence, specialized systems have been proposed for accessing such new kinds of databases, containing nonconventional data, such as images, videos, temporal series, maps, and so forth. Furthermore, the idea of information repository has been deeply influenced by the beginning of the Web age. Different visual systems have been proposed to cope with the need for extracting information residing on the Web. In particular, providing users with visual representations and intuitive user interfaces can significantly aid the understanding of the domains and knowledge represented by ontologies and linked data. As ontologies grow in size and complexity, the demand for comprehensive visualization and sophisticated interaction also increases. Ontology visualization and ontology visual querying are not new topics and a number of approaches have become available in recent years (see, e.g., Catarci et al., 2004; Soylu et al., 2015), with some being already well established, although more work is needed to provide the users with powerful querying and navigational aids and comprehensive visualization techniques.

Visual Querying of Data Streams

In the remaining of this chapter, we concentrate on a new class of VQSs specifically targeted for data stream processing that, despite its growing importance, has not been covered sufficiently in the literature.

As in classical VQLs, it provides a language, consisting of a set of visual constructs, to express, in a visual format, queries on data transmitted in a continuous and unbounded fashion (i.e., data streams). It is important to note that this class of languages can be seen as an extension of the generic visual query languages, due to the fact that they can query in a visual manner using the same criteria of classical VQLs both data stream and classical relational databases. They are oriented to a wide spectrum of users, even those who have some knowledge about concepts related to data streaming but with no skills in developing code.

The importance of data streams is continuously increasing. At the same time, all the existing VQLs working on classical databases are not good anymore for interacting with this huge and potentially infinite amount of data. Visual query languages for data streams are being developed to address these challenges. Basically, they use the same approach as the classical VQLs, extending it with the new data-stream operators. However, querying a data stream usually requires constructs different than a relational language. Due to the lack of a standard proposal, a large number of academic and commercial data stream query languages and their corresponding data stream management systems (DSMSs) have been developed. The main functionalities of these languages rely on some relational query language, typically SQL, which is extended to provide features dealing with characteristics present in streams, mainly temporal aspects, i.e., the fact that events change over time (e.g. *window* and *filtering operators*). Some examples of data stream query languages are SARI-SQL (Rozsnyai, Schiefer, & Roth, 2009), EP-SPARQL (Anicic, Fodor, Rudolph, & Stojanovic, 2011), CQL (Arasu, Babu, & Widom, 2003), Esper's EPL (http://www.espertech.com/esper/), SiddhiQL (http://docs.wso2.org/display/CEP300/Siddhi+Language+Specification) and StreamSQL.

Although there are a lot of data stream query languages, there are only few VQLs that are also able to interact with data streams. Only in recent years have the first VQLs for data streams appeared, either in the form of research prototypes (e.g., Bauleo et al., 2014) or real commercial systems. In particular, StreamSQL, along with its graphical counterpart EventFlow, is used within the StreamBase Complex Event Processing system (http://www.streambase.com), a commercialization of the Aurora project (Abadi et al., 2003). To date, this is the only existing commercial system.

Fundamentals

In the literature there are still a few systems offering visual query languages for data streams. What such a system should basically do is to apply the already existing concepts of VQLs for classical databases to new data-stream operators. In general, all visual query systems have a graphical development environment, usually consisting of (a) a canvas, where the visual constructs are freely arranged in order to compose the final query, and (b) different tabs containing the items necessary to produce the query visually.

StreamBase Studio is the graphical development environment of StreamBase. Here, users can graphically build a StreamBase application using a visual query language called EventFlow. A query is composed by connecting to each other the visual operators of the language and setting their properties. The palette with the visual elements is organized into three different tabs, as shown in Figure 24.8: (a) operators and adapters (such as filters and generic stream functions), (b) data constructs (such as a JDBC table to store information used by associated StreamBase operators), and (c) streams (e.g., input and output streams as well as an error stream).

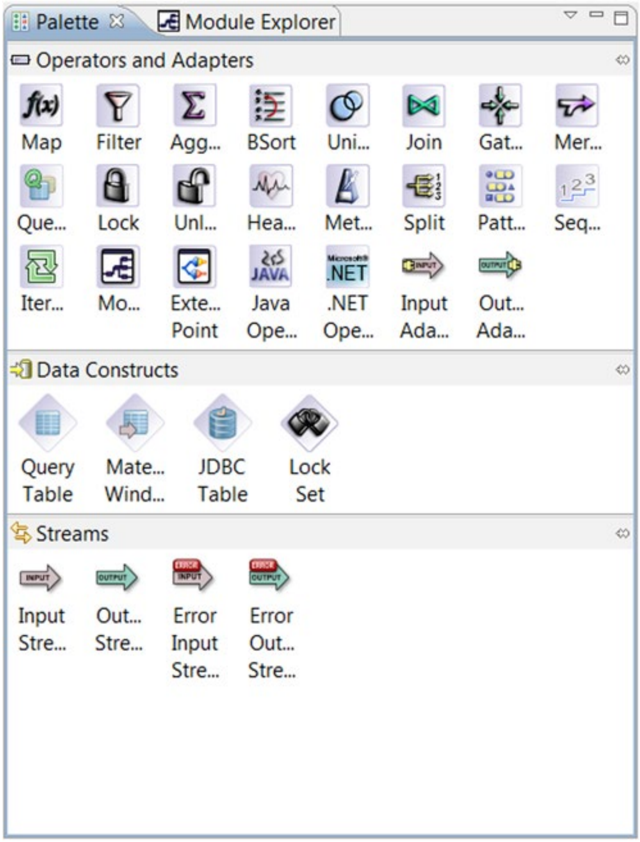


Figure 24.8 StreamBase visual operators.

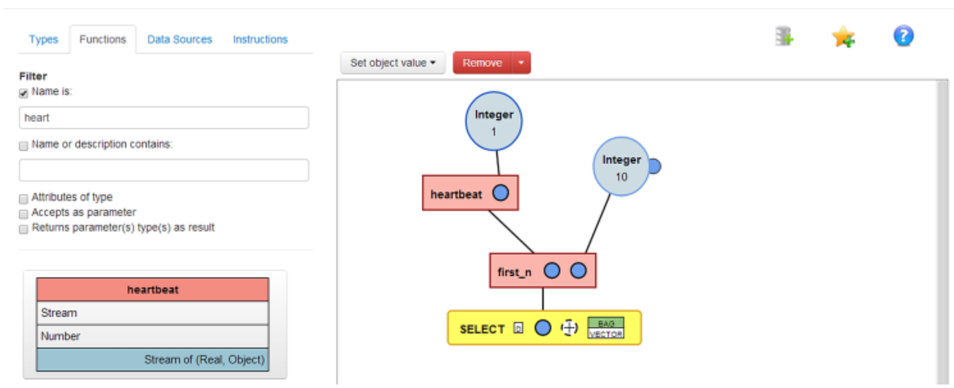


Figure 24.9 The visual query editor of the SmartVortex Visual Query System.

In the same way, the SmartVortex Visual Query System (Bauleo et al., 2014) has a powerful Visual Query Editor (VQE) containing the canvas and four tabs with all the necessary visual constructs to build a query. The VQE can be seen in Figure 24.9. The VQL is built on top of a federated data stream management system and generates its object-relational and functional query language. Therefore, the four tabs of the VQE contain all the types and functions of the textual language along with the instructions to develop the query. Since the system is intended for use in industrial applications, the last tab is specific to this context, containing all the machines/sensors managed by the system (i.e., the input streams). The four tabs along with an example of visual query are shown in Figure 24.9. For further details, the reader can consult Bauleo et al. (2014).

Having introduced these two systems, it is appropriate to note how the philosophy in the query-building process is slightly different. Whilst in the SmartVortex Visual Query System a visual query is completely built connecting the different items of the VQL, in StreamBase end users are forced to mix the graphical language with some snippet of code (e.g., setting the filter expression of a filter operator, setting the properties of a window operator, etc.). In both systems, the creation of a visual query requires some basic understanding of database and data-stream notions. A comparative user evaluation between these two systems is presented later.

Key applications

VQLs for data streams have been primarily proposed to allow nonprogrammers to express queries on data streams without concerning themselves about syntax and development code.

Key applications are related to each application area involving the processing of huge amounts of data. Financial services, air traffic control systems, industrial applications are all good examples. These kinds of applications are increasingly important mainly because these activities are increasingly becoming part of everyday life.

In Bauleo et al. (2014) it is possible to find some use cases taken from the industrial domain. In particular, two real industrial scenarios are presented in which the depicted VQS is used to monitor and analyze raw sensor-reading data, coming from industrial equipment, to figure out whether they behave as expected or not. Through the development of specific visual queries, running on the data streams produced, industrial engineers can solve common problems such as: (a) analyzing and predicting system/product performance (including its maintenance, fault detection, fault prevention, fault diagnosis), (b) analyzing and predicting the usage of system/product/service, and (c) analyzing and predicting customer service/product needs.

To facilitate the data stream processing operations and to make this task within the reach of many different types of users and needs, many other applications must be developed (both desktop and mobile), taking into account the new emerging multimodal interfaces such as touch, voice, and gestures.

Evaluation

Being a new research field, there are only a few studies that compare the usefulness of a specific VQL for data streams with respect to the equivalent textual data-stream query language.

A thorough study is reported in Bauleo et al. (2014), where a set of users evaluated the system in terms of effectiveness, user satisfaction, and efficiency. End users used the system, implementing two visual queries to solve problems related to industrial equipment. The realized queries took data from both a log file and a data stream produced from a machine.

The effectiveness was measured, in terms of (a) ease of access to the main features of the application, (b) ease of use, (c) ease of correctly understanding/interpreting the visual query language constructs and the visualizations of the query results, and (d) usefulness of the different functionalities provided by the system. Effectiveness and user satisfaction (i.e., how comfortable the users feel using the system) measures were collected by having participants respond to specific and relevant questions in a questionnaire. Efficiency was measured in terms of general user perception and the actual time spent to complete a query (to be more precise). The results were generally positive. Specifically, only 13% of users found it difficult to understand correctly/interpret the meaning of the visual query constructs; 12.5% of users found the system slow. There were no evident negative results for the other experimental measures.

The study also entailed a comparative user evaluation between the Smart Vortex visual query system (Bauleo et al., 2014) and the StreamBase application mentioned above. Results showed that, in general, the Smart Vortex visual query system outperformed the other system with the exception of the efficiency measures, where there were different outcomes for the two systems, depending on the kind of visual query.

Concluding Remarks

The current landscape is pushing information and data management at the center of any technical innovation, not only in computer science but in almost all fields. Big data, streaming data, Internet-of-Things, etc. are the recent trends that ultimately require users (without programming skills but some knowledge of data management, as data scientists have) to be able to understand the datastore content, focusing on meaningful items, finding query patterns, and reasoning on the query result. In this chapter, we have reviewed many years of research and innovation in Visual Query Systems, i.e., systems that use a visual representation for both the domain of interest and requests related to the domain, to provide access to the information contained in a datastore. We have then focused on the field of data stream processing, in order to show how previous approaches and techniques can be of inspiration in order to be merged in novel ways for addressing the new challenges of data management that we will be facing in the coming years.

Acknowledgements

This work has been partly supported by the EU Commission through the FP7 SmartVortex project. The authors would like to thank the many people involved in the design and development of the SmartVortex VQS, namely Mariano Leva, Alessandro Russo, Francesco Lanunziata, Arlind Tole, Emanuela Bauleo, Serena Carnevale, Donato Summa, Alberto Malagoli, Sonia Bergamaschi, Lars Melander, Tore Risch, and Cheng Xu, from Uppsala Universitet, Sweden for all the help and support during the use of their data stream management system.

References

Abadi, D. J., Carney, D., Çetintemel, U., Cherniack, M., Convey, C., Lee, S.,… Zdonik, S. (2003). Aurora: A new model and architecture for data stream management. *VLDB Journal*, *12*(2), 120–139.

Angelaccio, M., Catarci, T., & Santucci, G. (1990). QBD: A graphical query language with recursion. *IEEE Transactions on Software Engineering*, *16*, 1150–1163.

Anicic, D. Fodor, P., Rudolph, S., & Stojanovic, N. (2011). EP-SPARQL: A unified language for event processing and stream reasoning. In *Proceedings of the 20th International Conference on the World Wide Web (WWW 2011)* (pp. 635–644). doi:10.1145/1963405.1963495

Arasu, A., Babu, S., & Widom, J. (2003). *The CQL continuous query language: Semantic foundations and query execution* (Report No. 67-2003). Retrieved from Stanford InfoLab website http://ilpubs.stanford.edu:8090/758

Badre, A., Catarci, T., Massari, A., & Santucci, G. (1996). Comparative ease of use of a diagrammatic vs. an iconic query language. *Interfaces to Databases. Electronic Series Workshop in Computing*. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.45.2615&rep=rep1&type=pdf

Balkir, N. H., Ozsoyoglu, G., & Ozsoyoglu, Z. M. (2002). A graphical query language: VISUAL and its query processing. *IEEE Transactions on Knowledge and Data Engineering*, *14*(5), 955–978.

Bauleo, E., Carnevale, S., Catarci, T., Kimani, S., Leva, M., & Mecella, M. (2014). Design, realization and user evaluation of the SmartVortex visual query system for accessing data stream in industrial engineering applications. *Journal of Visual Languages and Computing*, *25*(5), 577–601.

Catarci, T., Costabile, M. F., Levialdi, S., Batini, C. (1997). Visual query systems for databases: A survey. *Journal of Visual Languages & Computing*, *8*(2), 215–260.

Catarci, T., Dongilli, P., Di Mascio, T., Franconi, E., Santucci, G., & Tessaris, S. (2004). An ontology based visual tool for query formulation support. In *Proceedings of the 16th European Conference on Artificial Intelligence (ECAI 2004)* (pp. 308–312). Amsterdam: IOS Press.

Hvoreckýa, J., Drlíkb, M., & Munk, M. (2010). The effect of visual query languages on the improvement of information retrieval skills. *Procedia Social and Behavioral Sciences*, *2*, 717–723. doi:10.1016/j.sbspro.2010.03.090

Massari, A., Pavani, S., Saladini, L., & Chrysanthis, P. K. (1995). QBI: Query by icons. In *Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data* (p. 477). doi:10.1145/223784.223890

Morris, A. J., Abdelmoty, A. I., & El-Geresy B. A. (2002). Design and implementation of a visual query language for spatial databases. In *Proceedings of the Sixth International Working Conference on Advanced Visual Interfaces (AVI 2002)* (pp. 359–360). New York, NY: ACM.

Murray, N., Paton, N., & Goble, C. (1998). Kaleidoquery: a visual query language for object databases. *Proceedings of the International Working Conference on Advanced Visual Interfaces (AVI 1998)* (pp. 247–257). New York, NY: ACM.

Polyviou, S., Evripidou, P., & Samaras, G. (2004, June 28–29). *Query by browsing: A visual query language based on the relational model and the desktop user interface paradigm*. Paper presented at the Third Hellenic Symposium on Data Management, Athens, Greece.

Reisner, P. (1988). Query languages. In M. Helander (Ed.), *Handbook of human-computer interaction* (pp. 257–280). Amsterdam: North-Holland.

Reiss, S. P. (2002). A visual query language for software visualization. *Proceedings of the IEEE 2002 Symposia on Human Centric Computing Languages and Environments* (pp. 80–82). doi:10.1109/HCC.2002.1046353

Rozsnyai, S., Schiefer, J., & Roth, H. (2009). SARI-SQL: Event query language for event analysis. *Proceedings of the IEEE Conference on Commerce and Enterprise Computing (CEC '09)* (pp. 24–32). doi:10.1109/CEC.2009.14

Soylu, A., Kharlamov, E., Zheleznyakov, D., Jiménez-Ruiz, E., Giese, M., & Horrocks, I. (2015). OptiqueVQS: Ontology-based visual querying. *Proceedings of the International Workshop on Visualizations and User Interfaces for Ontologies and Linked Data (VOILA '15)* (pp. 91–96). Aachen: CEUR.

Ziegler, J. E., & Fahnrich, K. P. (1988). Direct manipulation. In M. Helander (Ed.), *Handbook of human-computer interaction* (pp. 123–133). Amsterdam: North Holland.

Zloof, M. M. (1997). Query-by-example: A database language. *IBM Systems Journal*, *16*(4), 324–343.

25

Interfaces for Music

Ben Challis

Scenario

You have been tasked with devising and running a music workshop for adults who have a broad range of physical challenges. The group you will be working with is keen to make music, record, and perform together—they want to jam ideas out, create loops and layer textures and beats. Some of the group have previous experience with making music but can no longer manage the intricacies of their favored instrument. Translating ideas into actions is much more demanding than it used to be; finger dexterity is limited and then there is general fatigue to contend with. Some of the group have not been able to learn to play a traditional musical instrument so this will be their first time making music in a group like this. You head for the session armed with a range of hand percussion and several choice digital musical instruments; one of them is your phone. You choose one instrument because it transforms small finger movements into larger musical gestures; beats and grooves can be triggered and shaped with minimal dexterity using only a touch-pad interface. You choose another because it uses a grid of buttons to enable the deliberate building of loops and harmonic patterns—the performer can achieve rhythmic precision but in an unhurried and methodical way. On your phone you will run an app that maps hand position and orientation onto scales and chords—the performer can improvise in a key and scale that fits with the other instruments. The hand percussion can be used to build up rich rhythmical layers to underpin the main beats. Your workshop will be inclusive at all levels with everyone being able to contribute to an expressive collective musical performance.

Introduction

The last few decades have seen numerous technological advancements that have seemingly revolutionized the way in which we make, record, distribute, and consume music. There is no doubting that the emergence of digital approaches to music making and recording have been influential within this and, although there are clearly

The Wiley Handbook of Human Computer Interaction, Volume 2, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

specialist skills required to write, perform, and commercially produce music, there are many areas where new technologies are enabling us to achieve high-quality output with only limited formal training. Where once the multitrack recording studio was regarded as the exclusive domain of the professional sound engineer, it is now possible for high-quality recordings to be achieved using a fairly standard laptop computer, if not a tablet or smartphone. A typical digital audio workstation (DAW) will offer a virtual approach to multitrack recording (see Figure 25.1.), where the upper limit for audio tracks will only really depend on available processing power. Of course, achieving a truly professional finish still requires acoustically treated spaces but the essential elements of mixing-desk and off-board effects processors are now particularly accessible in terms of affordability but also usability. Eno (2007) refers to the recording studio as being a "compositional tool" in recognition of the creativity that is still being explored within the recording and production processes, and this concept is now even more firmly underlined within many common DAWs with the inclusion of tools and terminologies that reflect an array of musical processes that go beyond recording to include score writing, arranging, and, in the case of Ableton Live, for example, performance and even improvisation.

From the music performer's perspective, there are equally exciting opportunities to be realized. Digital musical instruments (DMIs) offer access to a vast array of sounds and musical textures, ranging from the emulation of acoustic instruments and historic electronic instruments through to the generation of new and unusual synthesized sounds. Not forgetting, of course, the augmentation of traditional acoustic instruments with digital technologies to offer new performance possibilities by way of hyperinstruments. There is something highly liberating about being able to access any number of different instruments at a moment's notice, making spontaneous creative judgements and choices along the way, working around that most recent idea again

Figure 25.1 The main window for Apple's Logic Pro X illustrates a fairly typical layout for a DAW featuring a horizontal timeline arrangement window with a mixing area populated with virtual faders.

but maybe using a djembe in place of a tabla or swiftly moving from a traditional flute to a shakuhachi and then to a soprano saxophone. When conventional instruments cease to take the music where we want to go there is still a seemingly endless variety of unusual sounds available to explore and, when the mood takes us, we can always draw upon tools that will allow us to create virtuosic patterns and textures that sweep across the underlying harmonic structures.

As with DAWs, there are significant gains to be made here in terms of affordability, portability and accessibility but there are numerous subtleties at play in terms of how we interact with all of these wonderfully creative tools. Given that this chapter is exploring the design of interfaces for music it might be anticipated that there will be a concluding section where a set of design principles are unveiled, which, if adhered to, will lead to "good" interfaces for music making. There are complications here, though, given the fundamentally creative and expressive nature of music, particularly where performance and improvisation are concerned. In terms of HCI, how can an interface for musical control be deemed to be "good" when typical metrics such as accuracy and speed may not always reflect the creative needs of the individual? By way of example, one performer's sense of "good" might reflect ease of rhythmical precision such that having notes artificially synchronized to a beat is of great benefit. In contrast, another performer may regard "good" as being a true reflection of her or his expressive intent—i.e. if the note is behind or ahead of the beat that is okay; learning to place notes accurately is part of the challenge of learning to play music. So, where HCI might ordinarily be concerned with task-based evaluation within interface design, for music performance systems there is likely to be greater emphasis being placed on the individual's overall musical experience.

A DMI might appear to be very difficult to work with when trying to achieve a specific expressive and meaningful musical outcome (e.g. the eigenharp—see Figure 25.2.). Perhaps there are multiple parameters being controlled within an interface that is occasionally awkward to manipulate. Maybe the final result is slightly off each time, a little behind the beat, wavering in pitch or inconsistent in tonal color. Maybe this is all acceptable if the end user is enjoying the challenge of mastering the instrument at a skill-based level. In one performer's context, maybe a "good" interface will allow failure just as much as it does success, maybe it will leave room for some sense of skillful mastery.

In contrast, maybe a DMI needs to be particularly easy to perform with. Perhaps the intended user has limited dexterity or restricted limb movement. Perhaps there are still multiple parameters being controlled but the performance context is within an ensemble such that certain basic musical elements need to be achieved to keep the performance together (e.g. Challis, 2013; Challis & Smith, 2012).The Kaossilator Pro (see Figure 25.3.) is a good example of a mainstream DMI that offers considerable musical variety while also being particularly accessible. Maybe in this context the desire to keep the music coherent has to outweigh some aspect of the individual's need for absolute control.

As with certain other areas in HCI, many of the factors that will influence good design practice will also be driven by individual needs such that any sense of generalization can prove difficult. This is particularly apparent when considering music and disability where interface design may be driven wholly by the needs of the individual's specific challenges. However, there are similar issues and compromises that are equally difficult to generalize upon when considering how to design interfaces and music

Figure 25.2 The Eigenharp, a particularly expressive DMI that requires considerable skill.

Figure 25.3 Korg Kaossilator Pro.

systems that enable nontraditional music learners to access potentially complex musical ideas or to enable highly experienced musicians to transfer their knowledge to unfamiliar instruments and new musical territories.

This chapter attempts to cast some light on what good interface design might entail within a context of accessible music interaction by identifying a set of music performance considerations to consider within the design process. As a preamble to that main discussion, the following section presents some opening thoughts on the breadth of music performance.

Music Performance

Although we can safely assume that the primary reason of existence for any given DMI will typically be music making, the context in which this creative practice is likely to occur may dictate the way in which the instrument should behave. This will ultimately influence the design of the interface by which the performer will interact with and control the flow of musical ideas. For many people, the notion of performance in music is likely to suggest one or more musicians creating music for a live audience. However this is perhaps a rather narrow perspective given that the same skills required to achieve a live performance will be in action during a recorded performance in a studio. The presence of an audience who will respond to the aesthetic appeal of a performance, reacting appropriately to the virtuosic and the inept or the expressive and the lifeless, does bring an additional relationship into the overall experience. This is in contrast to the recorded performance where mistakes can be rectified and multiple takes can be used to enhance the final experience. So, a musical performance can have an audience but does not necessarily require one; a novice musician rehearsing in private is still performing music but the judgement by which the quality of the performance is measured will be internal rather than external.

The music being performed might be composed or arranged and though offering room for expressive interpretation will often be following cultural rules and codes that dictate genre and style. Even where music is being improvised, these same conventions might still have place, keeping the improvisation idiomatic in nature by sounding "bluesy" or suggesting an off-beat reggae feel or drawing on hip-hop beats and grooves for example. Ultimately, there is free improvisation where the music is realized, performed and perhaps even lost within the moment, abstract in nature, free from predefined structure yet not without expression or the potential to draw upon the familiar. For much of the following discussion, the focus will be placed on highlighting issues that emerge as DMIs are designed or enhanced with a view to becoming more readily accessible as the same features that make an instrument appear to be enabling in one performance context might actually be quite limiting in another.

Performance Behaviors

Perhaps the most significant difference that becomes apparent when considering musical performance using DMIs is the ability for the underlying system to output far more information than the initial input might suggest. Consider a common two-button computer mouse as an interface to a software musical instrument. A simple performance opportunity could be devised by mapping the mouse movement and button presses to common musical parameters. For example, the horizontal position of the mouse cursor onscreen might dictate pitch, with volume being dictated by vertical positioning. These are commonly employed musical mappings and feature in a later discussion but for now let us also map the left mouse button such that sound is only ever produced if the button is down. In terms of instrumental skill, this simple DMI shares certain performance characteristics that can be found in many acoustic instruments and in terms of sound it may turn out to be reminiscent of a Theremin; one of the earliest electronic musical instruments. Invented by Leon Theremin (1925) (U.S. Patent Application No. 1661058) in the 1920s, this instrument (see Figure 25.4.) is noncontact in nature with physical gestures being translated into a continuous tone similar in sound to a bowed instrument. As with the Theremin, the performer using the mouse-based interface is very much in control of certain key aspects of the sound (pitch, volume and duration) and with subtle variations in either axis can achieve additional expressive qualities such as tremolo and vibrato. There it is, a simple interface that maps a performer's actions onto relatively intuitive musical parameters (see Figure 25.5.) whilst still requiring considerable skill to master. For ease of reference, let us give this new instrument a name: Moustrement.

Now consider this same interface but where the computer screen has been divided into a grid; notes from scales can be accessed horizontally and specific volume levels can be accessed vertically. The Moustrement still requires some skill to play but the pitches are being quantized to make more secure musical sense within key. The tremolo and vibrato effect could still be achievable with some allowance within the system and, overall, our revised instrument is starting to seem more accessible than the initial design. In terms of musical flexibility, the original idea has now evolved into a diatonic instrument (see Figure 25.6.).

For the next modification, imagine that, instead of a single note being produced, it is a three- or four-note chord that is sounded and that the tonality of the chord is matched to its position within the given scale or key (see Figure 25.7.). This is an interesting addition as the system is now introducing extra musical value into the output, a set of rules is being applied that calculates the correct notes for each chord

Figure 25.4 Leon Theremin performing with a Theremin.

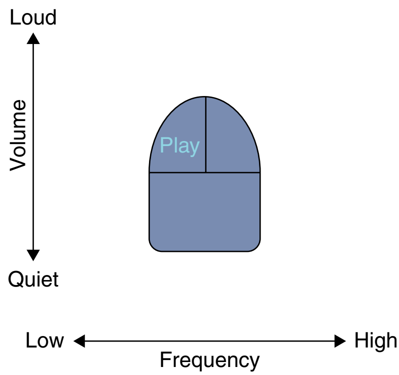


Figure 25.5 Musical mapping for the simple Moustrement.

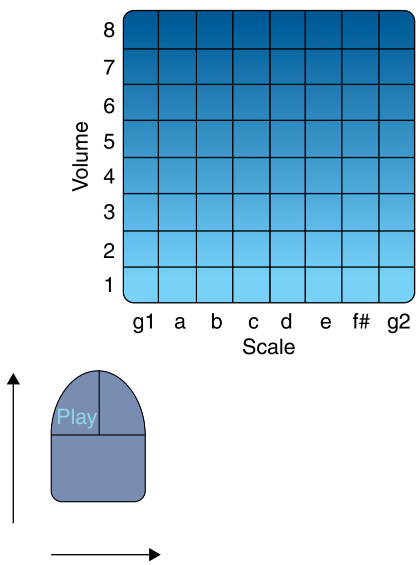


Figure 25.6 Revised musical mapping for the Moustrement. In this example, the only notes available belong to the scale of G Major.

based on a home key. Someone who has never played a musical instrument before could easily improvise a chord progression within a key that some traditional instrumentalists might regard as challenging. This is similar in principle to applying a capo to a guitar and moving one set of chord shapes into a different area of the fingerboard but with an acoustic guitar the chord shapes still need to be learned. In our simple DMI, there is no need to learn the notes of scales or the patterns of notes within chords to make these underlying harmonies work. This has the potential to be quite an empowering instrument given the complexity of output that might be achieved with only minimal formal musical knowledge.

Let us now take the Moustrement another step further by attaching a function to the remaining button, which introduces an arpeggiator. (An arpeggiator function is a very common feature on many synthesizers enabling patterns of ascending and

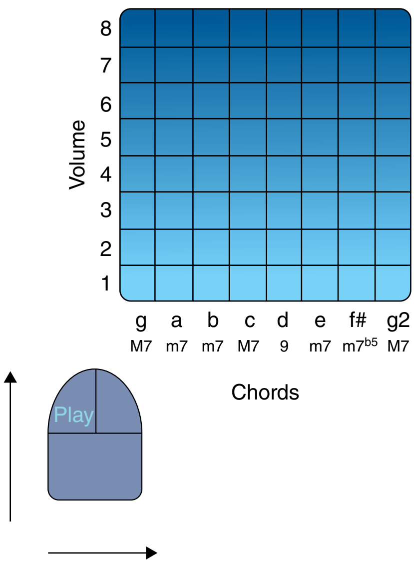


Figure 25.7 Moustrement revised to sound the chords of a given scale.

descending notes to be added onto a starting note or chord using a predefined tempo.) With no prior knowledge of chord construction, our novice performers can now race up and down broken chords within scales for which they may not even know the notes. How about if instead of arpeggiated chords the sequences being generated become pseudoimprovised phrases that match the underlying harmonies? As a next step, perhaps there is a bass line and drum pattern that follows the chosen harmonies beneath the improvised line, and so on…but where would all this additional musical output be coming from? The only significant musical choice being made is which chord to use yet there are now quite complete musical textures being produced. Our interface has not changed much at all yet the musical output has altered radically as has the performer's control over it. In essence, the interface is only one or two steps away from being a simple on-off switch to a precomposed piece. When so much of the control belongs within the system, is this really a performance still?

The underlying balance of control between user and system exists across other areas of HCI and for the purposes of music performance can be described in this context as an issue of musician versus machine. To fully appreciate the relationships that exist it will be useful to have a model for differentiating between different types and complexities of musical output and Malloch, Birnbaum, Sinyor, and Wanderley (2006) present just such a model by suggesting a continuum of performance behaviors where more and more control is relinquished from performer to machine.

Skill-based performance behaviors

The earlier examples of the Theremin and also the preliminary design for the Moustrement both offer skill-based performance behaviors. The performer has fundamental control over key musical parameters required for expressive performance: pitch, volume, duration. With these levels of control come subtle variations, which, in the case of both instruments, can lead to both vibrato and tremolo. Though they are both essentially electronic musical instruments, the performance behaviors being offered are very much in line with many acoustic musical instruments and with more sophisticated DMIs these might further include control of tone and timbre.

Rule-based performance behaviors

In the subsequent revisions to the Moustrement, additional musical enhancements have been introduced that are not directly under the control of the performer. Pitches are being quantized within musical scales and chords are being created automatically. Though the interface remains the same, various musical rules are being applied to bring additional value to the output. The rules being applied become more and more complex until the point where chords are being enhanced into arpeggiated patterns with variable rhythms and these ultimately become partially defined pseudorandom improvised melodic ideas and so on. It is still a very simple interface but an interface to complex rule-based performance behaviors.

Model-based performance behaviors

At the opposite end of the spectrum to skill-based performance is model-based performance where the control available to the performer may be so limited that the musical output is totally preordained, only to be released rather than being interacted with by the user. A single switch might be an interface to such a performance behavior, triggering a complete musical piece within a music sequencer with the only actions available being "start" and "stop."

It should be possible to place any given music-making system somewhere along this continuum depending on the level of control the interface allows between performer and musical output, although there will be some uncertainty as to where the exact boundary between one performance behavior and its neighbor(s) might appear to lie. However, the continuum as just described does not allow for the blurring of boundaries between model and skill-based outcomes even though there are music performance systems where there is ambiguity in just this area. To illustrate this, if a single switch is used to trigger a note-by-note release of a complete sequenced musical piece, using E-Scape for example (Anderson, 1996), where would the performance behavior belong within the continuum? Though there is no flexibility in the final piece in terms of note order, the expressivity that can exist within the deliberation over where to rhythmically place each note is entirely at the control of the performer. Add to this the ability to control the volume of individual notes by the strength at which the switch is pressed and this becomes more expressive still. Perhaps another aspect of the interface allows some level of control over timbre or simple vibrato. The musical order is predetermined but the expressive control remains firmly with the performer and, perhaps crucially, the performer still has the potential to go wrong.

There are other examples of music making systems that are instrument-like and appear to offer combinations of performance behaviors that fall outside of the constraints of the linear relationship as offered by a continuum and a powerful example of just such an approach will be covered in the following section that explores the notion of ownership and interfaces for music. With all this in mind, a more flexible

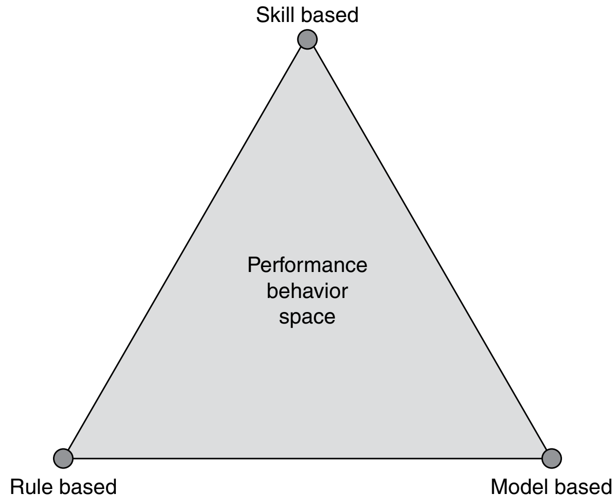


Figure 25.8 Enhanced model for placing DMIs within a two-dimensional performance behavior space.

Figure 25.9 Yamaha Tenori-on.

model might require a two-dimensional performance behavior space within which to place individual instruments; a triangular area, with each of the three performance behaviors being associated with its own corner. An instrument can be placed close to a single behavior without ruling out the possibility for pairs of behaviors to be in action if not potentially all three (see Figure 25.8.).

Other than allowing multiple instruments to be placed for comparison in a more flexible way, the space would also allow for instruments that offer multiple modes of interaction; the Yamaha Tenori-On for example (see Figure 25.9.). Individual buttons can be pressed to trigger notes in real time, and when used in playback mode,

Figure 25.10 Ableton Push controller.

step-sequenced musical patterns can be built up over time. Layers can then be edited, copied stored and manipulated in various ways including adding multiple layers into "blocks." These in turn can be brought together as larger pieces where all the parts to create a piece have been predefined ready for performance. This is true with DMIs like the Korg Kaossilator Pro (as mentioned earlier) and even more sophisticated composition, performance and improvisation tools such as Ableton Live using a Push controller (see Figure 25.10.). Pieces can be created ready for performance that might include further manipulation, almost certainly drawing upon rule-based behaviors but also touching upon skill-based as well.

Magnasson (2010) explores this same transition from micro to macro within music performance by enhancing a list of control modes initially proposed by Wanderley (2001) to include the following:

- *Filtering* (time and frequency domain manipulation of an audio signal).
- *Sonic texture generation* (layers of audio generated through synthesis or sampling).
- *Single musical notes* (where pitch, envelope, amplitude and timbre is controlled).
- *Continuous feature modulation of both note and phrase* (timbre, amplitude, pitch).
- *Musical gestures* (glissandi, trills, grace notes, etc.).
- *Simple scales and arpeggios* (of various speed, range, and articulation).
- *Phrases with different contours* (from monotonic to random).
- *Control of sampled material* (loop points, rate, granulation, pitch, filtering).
- *Synchronization of musical processes.*
- *High-level control of recorded material* (as seen with DJs).

In conclusion, Magnasson proposes an epistemic dimension space for comparing DMIs across eight axes addressing: expressive constraints, autonomy, music theory, explorability, required foreknowledge, improvisation, generality and creative-simulation. Magnasson's model enables the interaction requirements of any given DMIs to be further deconstructed in such a way that Malloch et al.'s (2006) generalized performance behaviors (skill, rule, model) can be more readily appreciated.

Ownership

Whilst still considering the musician v. machine issue, it is worth also considering the notion of ownership within this same relationship. Though we have considered how relatively simple interfaces can be mapped to potentially complex sonic outcomes it does not necessarily follow that this remapping of small gestures to big musical events carries as great an individual value as the challenge of producing events that may be seem less grand in comparison. The concept of musical "ownership" is something that Healey (2005) identifies as being particularly meaningful to the individual when considering the potential for assistive music technology in community settings. A given interface may offer great ease of use, enabling highly entertaining musical outcomes whilst also failing to allow the user much opportunity for developing a sense of creative contribution. In contrast, another interface might offer quite challenging access to a system that requires considerable skill to work with before achieving acceptable results. On the one hand an instrument that makes the performer appear to be particularly capable and, on the other, an instrument that can make the performer appear quite inept unless mastered over a period of time. If a sense of ownership holds significant value for the performer then it may be of no surprise if the latter experience transpires to be the most meaningful.

For an example of just how significant these individual values can be there is the story of Charlotte White to consider. At the age of 11, Charlotte received an injury to her head that resulted in the loss of all movement in her body. As a teenager, she was able to work with a music technologist (Doug Bott) at the Drake Music Project (UK). Bott explored different ways in which meaningful music performances could be achieved that would build upon Charlotte's childhood experiences of instrumental tuition. The music performance system that Bott ultimately designed enabled predefined chord shapes to be selected using switch-based access from one hand whilst individual notes from within the chord could be triggered by the performer making head movements using a magnetic motion sensor. Using this combination of simple head tracking and switch access led to Charlotte performing the Prelude to Bach's Cello Suite. Using the model for performance behaviors as described earlier, it is clear that skill-based behaviors are evident alongside rule-based (the defined chord shapes) within a suggestion of an overarching model-based behavior as the complete chord set is essentially the harmonic progression to one very particular piece.

There are video recordings of this performance available online (Drake Music, 2008) and the skill-based relationship between performer and interface is quite evident within these. The performance is not quite perfect with the rhythm momentarily losing its fluidity here and there but it is very much a performance where the audience can anticipate and relate to the playing of the note within the piece; in terms of ownership it is clearly Charlotte's performance.

Energy Exchange

In the Moustrement design example, the loudness of individual notes is dictated by the horizontal position of the mouse cursor; a fairly intuitive and quite common mapping of higher for louder, lower for quieter. Where an interface has no pressure sensitive component to work with, this approach to position-based mapping might well be practiced to a level where quite expressive results can be achieved. Consider the way in which loudness is normally determined within acoustic instruments, though. Although there are exceptions (harpsichord and church organ for example), the dynamic control that is so fundamental to expressivity is typically perceived in terms of an exchange of energy; the more vigorously an action is performed the greater the impact is on one or more aspects of the sound that is produced. Pressing, hitting, blowing, plucking, bowing—all of these performance interactions have a natural exchange of energy by which the instrumentalist will feel "connected" with the sounds being made.

This ability to anticipate change according to the level of energy required to perform a given action can be meaningful, ultimately enhancing the performance experience in terms of helping the instrumentalist feel directly connected with the musical output. It might even be that there is an expectation on the part of the user that adopting a more vigorous approach ought to affect the outcome in some way or other. There is evidence of this in some noncontact music systems, such as Soundbeam (Swingler, 1998) and Octonic (Challis, 2011), where performers can occasionally be observed attempting gestures that are not necessarily mapped to any additional musical parameter; shaking a hand or clenching a fist for example, where the only change actually being monitored is the distance from hand to sensor.

There may well be certain obvious actions that require some level of energy exchange that already map well within existing musical conventions (e.g. hitting pads, pressing keys, plucking strings), and these may be particularly meaningful within specific instrumental contexts. How about sounds that are less conventional though? Or where the performer does not have additional instrumental skill to draw upon or perhaps has additional challenges that make a conventional action less easily achievable?

"Good" interface design in some contexts may explore less conventional mappings of physical actions to musical outcomes that can still offer an appropriate sense of effort or exertion; squeezing, shaking, swiping, tapping, and so forth. Now would be a fitting point to momentarily return to the example Moustrement design. What opportunities for energy exchange might be included to make the connection between performer and dynamic range of our music interface more tangible? Using a mouse with velocity sensitive buttons could offer a good starting point or possibly a pressure sensitive pen of the type used in graphic art software. How about a simple breath-sensitive sensor, so that the user can blow to affect volume level, or a squeezable interface device to be held in a spare hand, or beneath a foot? A bowlike action might be achieved on a linear touch-sensitive resistor where a more vigorous swipe along the surface might correlate to a louder sound and so on. The point is that this relationship between energy input and musical output received can be important and should at least be considered within the design process. As with many of the other areas being touched upon, there are likely to be numerous opportunities that become available but that which feels "good" can only be determined by the end user based on individual needs and preferences.

Affordances, Constraints, and Musical Mappings

With acoustic instruments, the connection between performer and sound is so direct that the instrument can perhaps be perceived as an extension to the body. However, with a DMI, the performer is essentially decoupled from the sound production element such that the "instrument" can be considered as two components: an interface and a sound engine. Significant here is Norman's (1988) consideration of affordances as being perceived rather than adhering to Gibson's (1977) suggestion of all actions that might be possible. This notion has become embedded in HCI and in interface design in particular with affordances in this context being those actions that are physically possible but only against a backdrop of the user's objectives, beliefs, and previous experience.

It is clearly worthwhile considering whether particular components within an unfamiliar interface suggest ways in which we might interact physically but it will be equally worthwhile to consider whether the user might anticipate common musical mappings to be in force (loudness, tempo, pitch, timbre, duration, and so on). Before discussing affordances within particularly novel interfaces what of interfaces that appear to be derived from familiar acoustic instruments? With this in mind, Wanderley (2001) suggests three ways in which to consider DMI controllers:

- As being instrument-like—with interfaces that resemble acoustic instruments but that are reliant on synthesis engines.
- As augmented—that offer new capabilities to existing acoustic instruments.
- As being alternate—using sensors and input devices directly with no reference to acoustic instruments.

If a DMI is instrument-like there will be cultural affordances apparent within the interface that suggest ways of interacting with the instrument. Besides having a strong clue within its name, the Beamz laserharp is reminiscent of an acoustic harp, with stringlike laser beams that suggest that sound will occur if a plucking gesture is made. The Eigenharp discussed earlier looks very much like a wind instrument, a bass clarinet or bassoon for example, such that the even someone with no prior experience of playing a wind instrument will still recognize that this interface is probably going to make sound by blowing into it somewhere. The Korg Wavedrum looks like an acoustic snare drum or possibly a tambourine and our musical awareness of other things that are "drumlike" suggests that "hitting" will probably be significant. Although all three examples are instrument-like and suggest initial ways of interacting with the interface there may well be expectations as to how the mappings between interface and the underlying sound engine will work. Such cultural familiarity will undoubtedly come with cultural constraints attached. If a DMI is instrument-like and "blowable," then the new user will be justified in making the added connection that blowing harder will make the sound (whatever that might be) louder. Awareness of other similar looking percussion instruments would suggest not only that hitting the Wavedrum harder will make a louder sound but that striking different areas on the surface will produce a different timbre; this is not the case in practice. Likewise, plucking the "string" on a Laserharp ought to be somehow mirrored in a change in loudness but the reality is that the instrument's virtual strings are simply not dynamic. So in designing interfaces that are to some extent instrument-like, attention should be offered to the additional expectations that will come into play. Such levels of familiarity may drastically reduce the learning curve for working with the new instrument but with the added caveat that, in certain contexts (improvisation or special needs education for example), barriers may arise in terms of unexpected outcomes, for example hitting a drum produces the sound of a bowed string, sounds a chord, triggers a sequence, and so forth.

Overholt (2012) acknowledges this need for further distinction by extending Wanderley's three controller types into additional subcategories that can accommodate instrument-like controllers that employ alternate mappings to those that might be anticipated. Though Overholt's primary purpose was one of contextualizing an augmented acoustic violin that makes use of an embedded smartphone, the extended list of possible music controllers is an additional tool that can be drawn upon for classifying and comparing DMIs:

- Instrument-like controllers:
  - instrument-simulating controllers (mirroring playing techniques);
  - instrument-inspired controllers (abstractly derived techniques).
- Augmented controllers:
  - augmented by capturing traditional techniques;
  - augmented through extended techniques.
- Alternate controllers:
  - touch controllers (require physical contact with control surface);
  - noncontact controllers (free gestures—limited sensing range);
  - wearable controllers (performer always in sensing environment);
  - borrowed controllers (VR interfaces, gamepads, etc.).

In contrast to the example instrument-like controllers just described, how about those interfaces that can be identified as alternate? As identified earlier, there will be those physical elements that offer perceived affordances (pushing faders, turning knobs, pressing switches, swiping touchpads and so on) but perhaps of more immediate interest is whether there are commonly associated or anticipated musical mappings to allow for? There are certainly numerous cultural and semantic conventions that exist that will influence which musical outcome is perceived as being the most likely. For example, imagine having to move the volume dial on your media player in an *anticlockwise* direction to make the sound *louder* or that *lowering* the cut-off frequency on your synth's low-pass filter requires a fader to be pushed *upwards*. Armed with experience of many similar actions with all kinds of contrasting technologies it is reasonable to assume that movement from left to right or down to up will result in some level of increase and vice versa and there are many sonic and musical parameters that will map well to these same actions such as volume, pitch, decay, tempo etc.

As an example of a relatively common alternate controller, consider again Korg's Kaossilator Pro. Though there are faders and knobs that offer control over various parameters, the bulk of the performance control happens using a single finger on a small touch-sensitive surface. As described earlier whilst considering performance behaviors, this particular DMI allows relatively skill-based interaction where single notes are activated but with additional access to more complex rule-based behaviors such that patterns and beats can be triggered and manipulated. Some of the mappings against the available two-dimensional movement are relatively abstract and require learning (altered drum patterns for example) but others make use of the musical conventions that were just identified. A common effect to have available within a given sound patch is a low-pass filter and the low-to-high mapping for the cutoff point is matched intuitively to vertical movement. Similarly, individual pitches where available will map from left to right.

Multiparametric Control

With many acoustic instruments, the player is controlling more than just the volume and pitch of individual notes. For example, and perhaps most notably, the tonal color of wind instruments can be affected greatly by subtle changes to the embouchure. Similarly, the location of a bow or pick on a string, or beater on a drum skin will achieve comparable tonal contrast. In considering the design of new and novel interfaces for music performance it is worth considering how effectively the interface might offer multiparametric approaches to sound production and manipulation. There may be a tendency to try and deconstruct sound and music tasks into separate control streams whereby multiple parameters are then given over for precise but individual control. In practice, however, there is evidence to suggest that this method of individual parameterization may not be as effective for managing more complex sound control tasks akin to those found within acoustic instruments.

Using three simple comparison interfaces across a prolonged series of testing, Hunt and Kirk (2000) showed how even a simple interface to multiparametric control can be highly effective at enabling a user to engage with seemingly complex musical tasks. Using a strategy of "listen and copy" three test interfaces were considered. The first allowed individual control of four key sonic parameters using a mouse and onscreen graphical faders and the second used a group of dedicated hardware faders using one-to-one mappings. The final interface used a mouse and two faders to enable a series of many-to-one mappings such that the key sonic parameters were ultimately controlled through simultaneous interaction with most elements of the interface. The multiparametric mapping may have presented the most challenging interaction yet positive results with this approach greatly outweighed those with the more direct mapping and, significantly, many of the participants commented that there was something more engaging and entertaining about the multiparametric approach that perhaps encouraged or enabled elements of spatial thinking and how physical gestures might translate into sonic outcomes.

Functionality, Immediacy, and Improvisation

Much of the discussion thus far has focused on the issues and challenges to interface design for music performance but, as was identified in an earlier section, this encompasses a broad range of contexts and included within these is improvised music along with all the opportunities for spontaneity that can exist within performance. Again, there is a range of improvisational contexts that we might consider (musical play, music therapy, community music, the soloist, improvising ensembles and so on), which will, in turn, draw upon approaches and techniques that will exist across a spectrum that extends from idiomatic to abstract (or free) improvisation. One of the key aspects of improvisation with electronic instruments is the flexibility of moving swiftly and seamlessly between contrasting sounds and textures. This can have a dramatic effect on the ability of the performer to respond to the ideas that are emerging around them whether by reflecting back similar phrases and textures or by introducing radical change. One of the key barriers to improvisation within any given DMI is the immediacy of offering instinctive ways of altering the sounds in some way. There are many tradeoffs to be made here if an interface is to enable a performer to quickly access, modify, and generally interact with sound in a creative and intuitive fashion part way through an improvised performance.

Consider music therapy as an example music performance context. The therapist will employ technical and expressive performance skills to elicit reaction and response from individuals within a clinical setting, using improvisation as a means for suggesting dialogue, perhaps mirroring physical actions as musical gestures and in doing so establishing the basis for connection and communication. Magee (2006) identifies how the Theremin, as described earlier, lends itself to this kind of relationship being readily accessible by anyone whilst also offering advanced technique with some level of mastery. If the therapist were to use a DMI in place of an acoustic instrument within a particular session the instrument would still need to offer a high level of autonomy, allowing the spontaneous creation and manipulation of fresh ideas in a controlled and expressive way. In a different context, a DMI that is perhaps quite empowering for a less experienced player that would enable a child to play in a school ensemble, for example, might help minimize the margins for error, encouraging accurate playing of notes and phrases perhaps even ensuring that the music stays within a predetermined style.

Liveness and the Aesthetics of Performance

Music performance is often a shared experience between performers working together but also between performers and their audiences. In electronic music, the relationship between performer and audience can be troubled by an apparent lack of physical gesture to accompany the sounds the audience will hear. The use of high-level mappings or less visible interfaces can ultimately confuse audiences who subsequently find it harder to read and interpret the performers' actions (Paradiso, 1999). There are subtleties here whereby an audience will feel connected with the performance as the visual cues suggest likely outcomes; a sense of anticipation of the music to come. Though the notion of liveness and performance exists in other areas in the arts, it is easy to appreciate why the use of certain types of technology in live music performance might instill heated debates on the experiences, relationships and barriers that affect an audience's perception over whether any given performance "feels" live; having the performer present is not the only issue here.

Of all the technological candidates that deserve scrutiny within a context of music interfaces and liveness, the laptop must rank quite high. Brown, Bell, and Parkinson (2014) describe how "the performer commands control over a powerful, layered, mesh of sound, but as far as the observer is concerned, they might as well be checking their email." Blain (2013) identifies how the laptop has enabled the electronic musician to bring work out of the recording studio and into the arena of live performance arguing that, although there is a growing body of academic discussion and research in areas such as composition, instrument design and music education there has been little debate within the area of performance theory. Key within this is Ostertag's (2002) suggestion that "the measure of a work of art is whether one can sense in it the presence of the artist's body" yet the continued popularity of live solo performances where this connection is less apparent (e.g. John Hopkins, Fourtet, Plastikman, etc.) suggests that "liveness" in that sense is not necessarily an issue for all audiences.

This particular debate goes beyond the scope of this chapter but an awareness of the potential issues should not be ignored within the overall design process; a highly usable and efficient performance interface may just not be that interesting to watch. In looking to offer a more meaningful connection between performer and audience, there are justifiable arguments for considering the overall aesthetic of the performance as a visual experience alongside the key HCI issues.

Repurposing

Much of this chapter has been dedicated to recognizing key considerations for interface design in enabling musical performance. Appreciating the performance needs that are in place should help inform the design of one or more appropriate methods of control and interaction; the performance needs will ultimately dictate the design. However, there is now an expanding community of hackers and DIY music enthusiasts who are repurposing existing technologies to create alternative interfaces for music performance. This repurposing of technologies to create new and novel interfaces has partly been fueled by the rapid emergence of exciting and highly affordable technologies from within the computer games industry (force-feedback joysticks, haptic mice, gamepads, Wii-Mote, Kinect, Leap Motion, Oculus Rift, etc.) but also by the emergence of software platforms by which these potential "instruments" can be easily mapped to sound and music (Pure Data, Max/Msp, AudioMulch, Ablteon Live, OpenFrameworks, Apollo Ensemble, etc.) In many ways, there is a reversal to the design process here as the available actions within the interface are not alterable without adaptation. Much of the process of repurposing is concerned with which musical mappings might be made available across an existing set of knobs, buttons, and sensors. Where performance needs are ordinarily dictating the interface design, when repurposing an interface the performance opportunities will be dictated by the limits of the device being used and the rules and mappings that might be applied. However, the model for performance behaviors, as defined earlier, still has place in terms of understanding the varying levels of control the performer might have within the system.

Summary

If there is one key point worth keeping in mind whilst considering the design of any new DMI it would be that the typical measurements of success within interface design do not necessarily transfer that well when considering interfaces for artistic expression. Where designers might ordinarily be concerned with metrics such as speed and accuracy and therefore employ task-based approaches to interface evaluation, with interfaces for artistic performance there will be highly individualized values to allow for. An interface that is apparently complex, perhaps with multiple parameters being controlled and even with a noticeably steep learning curve, could have considerable appeal over a more simplistic interface that inevitably shifts the balance of control more towards the system than the player. This does not mean that simple interfaces that relinquish considerable control from performer to system do not have place or purpose within a context of music performance but there will be an associated creative cost being paid. The less potential a system has for enabling skill-based performance behaviors, the less involved the performer can feel as the balance moves further away from improvisatory freedom and gradually closes in on mechanistic constraint; the ultimate cost being a complete erosion of any meaningful sense of artistic ownership.

References

Anderson, T. (1996). "Composability": Widening participation in music making for people with disabilities via music software and controller solutions. In *Proceedings of the ACM Conference on Assistive Technologies* (pp. 110–116). New York, NY: ACM.

Blain, M. (2013). Issues in instrumental design: the ontological problem (opportunity?) of "liveness" for a laptop ensemble. *Journal of Music, Technology and Education*, *6*(2), 191–206.

Brown, O., Bell, R., & Parkinson, A. (2014). Examining the perception of liveness and activity in laptop music: Listeners' inference about what the performer is doing from the audio alone. In *Proceedings of the International Conference on New Interfaces for Musical Expression* (pp. 13–18). New York, NY: ACM.

Challis, B. (2011). Octonic: An accessible electronic musical instrument. *Digital Creativity*, *22*(1), 1–12.

Challis, B. (2013). Assistive synchronised music performance. In *Arts and Technology, Third International Conference ArtsIT, Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering* (vol. 116, pp. 49–56). doi:10.1007/978-3-642-37982-6

Challis, B., & Smith, R. K. (2012). Assistive technology and performance behaviours in music improvisation. In *Proceedings of ArtsIT 2011, Arts and Technology, Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering* (vol. 101, pp. 63–70). doi:10.1007/978-3-642-33329-3

Drake Music (2008). Charlotte White plays Bach—Cello Suite No. 1 Prelude (abridged) [Video file]. Retrieved from http://www.drakemusic.org/video/charlotte-white-plays-bach-cello-suite-no-1-prelude-abridged/

Eno, B. (2007). The studio as compositional tool. In C. Cox & D. Warner (Eds.), *Audio culture, readings in modern music* (pp. 127–131). New York, NY: Continuum.

Gibson, J. J. (1977). The theory of affordances. In R. E. Shaw & J. Bransford (Eds.), *Perceiving acting and knowing*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Healey, R. (2005). New technologies and music making. In P. Moser & G. McKay (Eds.), *Community music: A handbook* (pp. 161–179). Lyme Regis: Russell House Publishing.

Hunt, A., & Kirk, R. (2000). Mapping strategies for musical performance. In M. Wanderley & M. Battier (Eds.), *Trends in gestural control of music*. Paris: Ircam Centre Pompidou.

Magee, W. (2006). Electronic technologies in clinical music therapy: a survey of practice and attitudes. *Technology and Disability*, *18*, 139–146.

Magnasson, T. (2010). An epistemic dimension space for musical devices. *Proceedings of the 2010 Conference on New Interfaces for Musical Expression* (pp. 43–46). New York, NY: ACM.

Malloch, J., Birnbaum, D., Sinyor, E., & Wanderley, M. (2006). Towards a new conceptual framework for digital musical instruments. *Proceedings of the Ninth International Conference on Digital Audio Effects* (pp. 49–52). Montreal: McGill University.

Norman, D. (1988). *The design of everyday things*. London: MIT Press.

Ostertag, B. (2002). Human bodies, computer music. *Leonardo Music Journal*, *12*, 11–14.

Overholt, D. (2012). Violin-related HCI: A taxonomy elicited by the musical interface technology design space. In *Arts and Technology. Lecture Notes of the Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering* (vol. 101, pp. 80–89). Berlin: Springer.

Paradiso, J. (1999). The brain opera technology: New instruments and gestural sensors for musical interaction and performance. *Journal of New Music Research*, *28*, 130–149.

Swingler, T. (1998) The invisible keyboard in the air: An overview of the educational, therapeutic and creative applications of the EMS Soundbeam™. In *Proceedings of Second European Conference on Disability, Virtual Reality and Associated Technologies* (pp. 253–259). Reading: University of Reading.

Theremin, L. S. (1925). *U.S. Patent Application No. 1661058.* Washington, DC: U.S. Patent and Trademark Office.

Wanderley, M. M. (2001). *Performer-instrument interaction: Applications to gestural control of music* (Unpublished doctoral thesis). University Pierre et Marie Curie—Paris VI, Paris.

Part VII
Interaction

26

Embodied Conversational Agents

Hung-Hsuan Huang

Introduction

Machines can work without rest while maintaining levels of accuracy and quality that could never be achieved by humans. For decades, artificial intelligence researchers have pursued the goal of building machines that can engage in conversation with humans at a level close to that of other humans. The fictional video, *Knowledge Navigator,* produced by Apple in 1987, was a good example of this idea. The term "embodied conversational agents" (ECAs) was first proposed and defined by Justine Cassell in the book *Embodied conversational agents* (Cassell, Sullivan, Prevost, & Churchill, 2000) as "computer interfaces that can hold up their end of the conversation, interfaces that realize conversational behaviors as a function of the demands of dialogue and also as a function of emotion, personality, and social conversation." ECAs are usually realized as lifelike characters in 3D computer graphics (CG) animation (hereafter ECAs) and are the subject of this chapter.

An ECA, sometimes called a *virtual agent* or a *virtual human* is often basically a simulation of a person. The research issues therefore include how to make the computer-driven graphical character look like a real human (realistic outlooks and movements, i.e. CG animation) and behave like a person (take humanlike decisions at the right times).

In face-to-face conversations, we humans not only use language but also fully utilize our bodies to communicate with interlocutors. We adjust the tone of our voice according to the context of conversation and our emotional state; we perform hand gestures (Kendon, 2004), change body postures to supplement speech, and monitor those expressed by the interlocutors at the same time. In order to achieve these conversational functions on a machine, sensors are required to perceive verbal and nonverbal status of the human communication partners, and actuators are required to realize the agents' intentions as in a way that can be perceived by humans. The difficulties not only come from what the agent can do but also from the subtle differences in the quality of their movements and their appearance, which may cause different interpretation by users. *Uncanny valley* is a hypothesis proposed by Mori in the 1970s (Mori, 1970). It assumes that the relationship between the human likeness in appearance of

The Wiley Handbook of Human Computer Interaction, Volume 2, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

an artifact and the intimacy from human toward it can be viewed as a valley-shape curve. That is, when the artifact looks more humanlike, intimacy with it becomes higher. However, it dramatically drops at some point where the artifact looks very humanlike but does not behave so humanlike. Intimacy increases again when the artifact's behavior is improved. The curve is, then, in the shape of a valley, with higher intimacy at both side but lower at the middle. This hypothesis is a popular research topic and has been investigated by many research groups in past decades, including Hiroshi Ishiguro's group, which is remarkable for building lifelike humanoid robots (Zlotowski et al., 2015).

Why ECA Instead of Communicational Robots?

Comparing ECAs with mechanical counterpart—humanoid communication robots—they have potential advantages in that they have greater freedom in their faces and bodies, less noise in actuation, and less limitations in the virtual environment where they are deployed. Figure 26.1 shows a virtual instructor of ballroom dance developed by our group (Huang, Uejo, Seki, Lee, & Kawagoe, 2012). In this system, the ballroom dance student can practice his / her dance and get the instructions from a virtual instructor. During the interaction, not only does the student dance but the virtual instructor also performs dance animations. These animation sequences are based on data from a professional dance instructor obtained with a motion capture device. The dance steps involve complex, fast, and subtle body movements, which are difficult to realize with a robot.

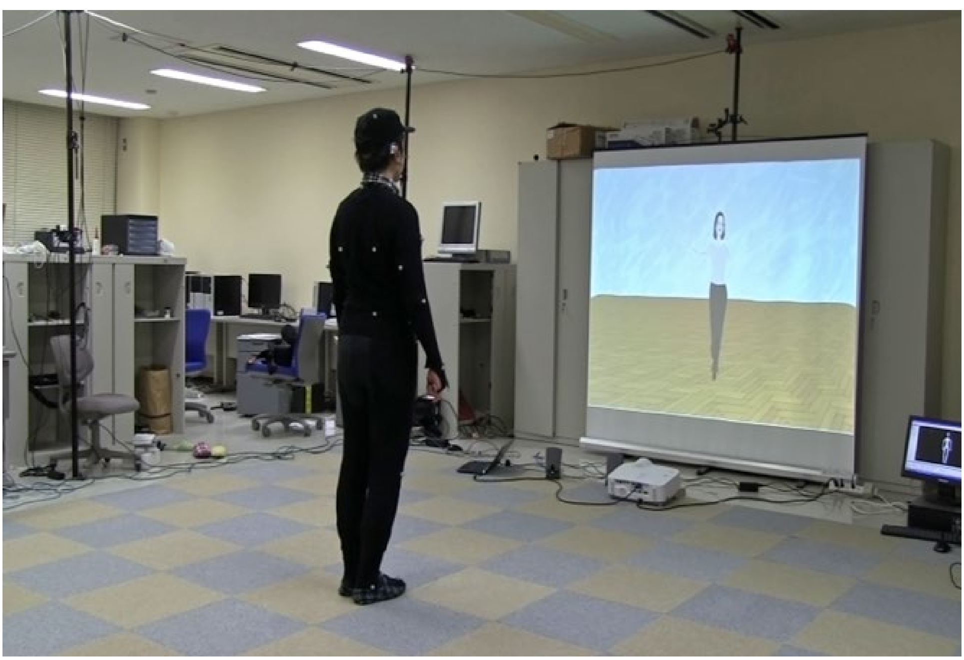


Figure 26.1 Virtual instructor ballroom dance system.

The ECA research projects share the same goal: the realization of humanlike behaviors of artifacts with communication robots. They have the same difficulty in perception processing and have similar logic in the decision-making module. Researchers also share research interests on both sides. Therefore, research papers related to robots are presented at agent-focused conferences like Intelligent Virtual Agents (IVA), Autonomous Agents and Multiagent Systems (AAMAS), and Human-Agent Interaction (HAI).

A summary of the advantages and disadvantages of ECAs and communication robots is listed in Table 26.1. Their inherent weakness is the lack of actuators: that they cannot have physical access to real-world objects. Despite that, ECAs relieve researchers from the mechanical and material issues associated with robots, with the relatively lower hurdle of rendering and animating computer graphic characters realistically. This allows them to concentrate on realizing high-level and advanced conversational abilities like speech-synchronized lip movements, and rich facial expressions with synchronized and sophistical movements involving all parts of the face. Embodied conversational agents are therefore considered as ideal interfaces for applications such as the simulations in psychology studies, language training, entertainment purposes, or public services, where high-level communication abilities are required.

Table 26.1 Comparison between ECAs and robots.

	ECA	*Robot*
Advantages	Low cost Can be implemented on handheld mobile devices Widespread commercial products Can be implemented in a fully immersive virtual reality environment with a HMD	Strong feeling of presence to improve the engagement of the user(s) in the interaction Can be equipped with actuators and have physical access to real-world objects Self-propelled according to the robot design
Disadvantages	Mona Lisa effect Weak feeling of presence to obtain the engagement of the users No actuator and therefore no physical access to real-world objects Stationary setup in most cases	Cost is so high that used for research projects and industrial purposes only Difficult to achieve realistic, humanlike facial expressions and body movements due to physical constrains
Suitable application domains	Pedagogical or training applications Commercial products marketed for general public consumers Psychological experiments Games	The applications where the responses to individual users is essential in a multiuser setting The applications where physical access to real-world objects or self-propelled mobility is required

The lower demand for resources in the development of ECAs is another advantage. Although the level of autonomy or expressiveness is probably lower from the researchers' viewpoint, the user interface with characters has been widely employed in commercial applications and Web sites. Another point that should be noted for ECAs is a phenomenon called the *Mona Lisa effect*. Many ECAs setups use a screen-projector combination or large-size displays, which render the characters in two dimensions. In a situation where multiple users are interacting with a system in this kind of setup, all people will perceive basically the same gaze and face direction from the characters. That is, when the character is looking forward, then all of the users feels that the character is looking at them; when the character is looking to the right, then each one of the users feels that the character is looking at someone to the left, and so on. This phenomenon may confuse users when the agent's behaviors are directed at only one person in the user group. In order to address this issue, our group proposed a setup to compensate the 2D agent with a physical pointing device in a quiz game application where the quiz-master agent interacts with multiple users (Huang et al., 2010).

To summarize, there is no absolute advantage for neither ECAs or communication robots when they are compared with each other. In some situations, an ECA setting is more appropriate (e.g. where the expressiveness of realistic human behaviors is essential) whereas in the other situations a robot is more appropriate (e.g. multiparty interaction or where access to physical objects is essential). There is also some work that combines the advantages of both of these setups. For example, the prototype of a robot bartender, JAMES (Gaschler, Kessler, Petrick, & Knoll, 2015), has a graphical face rendered on a tablet that is mounted on a robot body. All in all, user interface developers need to thoroughly consider what they can get from the setup to achieve best system performance.

Brief History of ECA Developments

With the advance of computer hardware, computer graphics, natural language processing, speech recognition, and synthesis technologies, ECAs have attracted great interest from researchers in the past two decades (Nishida, 2007; Prendinger & Ishizuka, 2004). A wide variety of embodied conversational agent systems have been developed by a number of research institutes. For example, Rea (Real Estate Agent) (Cassell et al., 1999; Cassell, Bickmore, Campbell, Vilhjálmsson, & Yan, 2000) is an ECA who mediates house information with a single user. Rea uses simple heuristics on verbal and nonverbal behaviors by the user to carry out conversational turn management; she yields the turn to the user when the user starts speaking and terminates her own utterance in the middle when the user starts to make gestures. She also generates synchronized multimodal utterances.

Media Lab Autonomous Conversational Kiosk (MACK) (Cassell et al., 2002; Nakano, Reinstein, Stocky, & Cassell, 2003) is an ECA who can answer questions about and give directions to the MIT Media Lab's research groups, projects, and people. MACK uses a combination of speech, gesture, and the indications on a map placed on a table between himself and individual users. The user's head movements and gaze directions are tracked by MACK so that he can estimate whether the user has understood what he just said (grounded) and decide whether to proceed or explained his actions in more detail.

Greta (Pelachaud, Carofiglio, de Rosis, & Poggi, 2002) is a doctor agent who gives her patients information about drug prescriptions. She is implemented as a 3D talking head with her own personality and social role, and the capability of expressing emotions consistent with the conversational context of her own goals.

Max (Multimodal Assembly eXpert) is a virtual human developed in Bielefeld University and is adopted as various roles with different abilities. As an assistant to the human user it can collaboratively construct virtual objects (Kopp, Jung, Lebmann, & Wachsmuth, 2003) with multimodal interaction, is a master of a card game with an emotion simulation (Becker, Prendinger, Ishizuka, & Wachsmuth, 2005; Boukricha, Becker, & Wachsmuth, 2007), and a science museum guide (Kopp, Allwood, Grammer, Ahlsen, & Stocksmeier, 2008; Kopp, Gesellensetter, Kramer, & Wachsmuth, 2005) with real-time feedback to visitors' keyboard inputs.

As noted in the introduction above, the earlier ECA work focused on the development of fundamental functions of ECAs—for example how to generate realistic animation of facial expressions and body movements and how to process dialogue with users, and so forth. In recent years, thanks to more refined CG animation technology and the emergence of smart mobile devices with sensors at low prices, ECA research has been shifting to its next stage. More and more projects explore human-agent interaction at deeper levels of communication. SimSensei (DeVault et al., 2014) is a virtual therapist platform developed for the use in the clinical and health-care fields. The system can automatically analyze psychological stress like depression, anxiety, and posttraumatic stress disorder (PTSD) in real time according to facial expressions, body postures, acoustic features, and linguistic features obtained from video/audio inputs. TARDIS (Training young Adult's Regulation of emotions and Development of social Interaction Skills) is an EU project aiming to provide a social skill training environment for job interviews (Jones et al., 2014).

Standardization of ECA Development

As a result of emerging research interests on virtual human animations and the demand for the standardization of ECAs, there are already a number of activities trying to standardize the production of CG characters or autonomous ECAs. In this section, we introduce them in two categories. First, activities attempting to propose a standard description language of character animations. Second, framework that is being developed that is meant to address the standardization of the behaviors expressed by autonomous ECAs.

Character animation description languages

Some high-level conversational agent or virtual human description markup languages have been proposed or are being developed, such as AML (Avatar Markup Language) (Kshirsagar et al., 2002), VHML (Virtual Human Markup Language) (Gustavsson et al., 2001), CML(Character Markup Language) (Arafa & Mamdani, 2003), APML (Affective Presentation Markup Language) (Carolis, Pelachaud, Poggi, I., & Steedman, 2004), and MURML (Multimodal Utterance Representation Markup Language) (Kranstedt, Kopp, & Wachsmuth, 2002). Avatar Markup Language is a high-level script language specifying avatar animations; the AML processor reads AML scripts containing high-level descriptions of avatar facial expressions, body animation, and utterances of the avatar, or references to MPEG-4 FBAP (Facial Body Animation Parameter) (Pandzic & Forchheimer, 2002) files, and then it generates the corresponding MPEG-4 bit stream for Web-based applications. However, the agent architecture is deterministic and thus has no flexibility; the script language does not consider the input part from the human user, either. Virtual Human Markup Language is a high-level markup language that describes a virtual human for general purposes; it is composed with a set of sublanguage includes descriptions of emotion, facial expressions, gestures, and so forth. However, the specification of VHML is distinct and thus has little flexibility to include supplement FAP/BAP files like AML does. Many parts of it are still undefined, especially the gesture or body animation parts. Character Markup Language is another under development high-level virtual character description markup language, which is similar to AML. It differs from AML in the specification of emotion and personality while its predefined base set of movements cannot be extended dynamically. Affective Presentation Markup Language is a language that specifies the association of verbal utterance, facial expression, and dialog moves (Traum et al., 1999) of a talking-head agent. Multimodal Utterance Representation Markup Language associates gestures with begin/end timing marks that are inserted into verbal utterances. Each gesture is described with a set of parameters presenting wrist location, hand shape, and wrist orientation.

Multimodal Presentation Markup Language (MPML) (Prendinger, Descamps, & Ishizuka, 2002), MPMLVR (Okazaki, Aya, Saeyor, & Ishizuka, 2002) and TVML (TV program Making Language) (Hayashi, Ueda, Kurihara, & Yasamura, 1999) are very high-level script languages designed to enable presentation or TV-program-like content to be made easily. With its user friendly interfaces, this content can be created by writing a simple script to describe a limited, predefined set of virtual words, objects, characters, and character behaviors.

Work on these languages has been intensive, but none of them have been widely accepted as a de facto standard. This may be due to the following reasons.

It is difficult to find a balanced and thorough coverage of a high-level description language. A virtually infinite number of possible behaviors can be carried out by both human and humanlike characters. Character animations that are considered to look natural vary from application to application, character to character. Therefore, in most cases, the languages can only be specified as extremely high level where concrete specifications can be figured out. This limits the benefits of adopting such a language rather than a home-made description language that is most suitable to the researchers' own purpose. The same reason also resulted in the fact that most of these languages are similar to each other but none dominates. There is a lack of a bundled compliant character animation toolkit. Most of the work mentioned above does not provide a fully functioning character animation toolkit, except MPML dialects and TVML. If a description language neither specifies the animations concretely nor provides an animation toolkit, it is hardly useful for ECA developers. On the other hand, although MPML and TVML provide easy-to-use and fully functioning toolkits, they cannot be extended easily, and thus their application is limited.

In contrast with the languages mentioned above, MPEG-4 FBAP is a specification that tries to achieve video communication between conversation partners and avatar animations through a narrow network channel. Detailed character animation parameters are specified in this standard, and the CG character is animated like a virtual robot, i.e. rotating the joints in the sense of angles. A VRML97 (Virtual Reality Modeling Language) (Web3D Consortium, 1997) based representation standard of humanoid model, H-Anim (http://www.hanim.org) is adopted. There are 66 low-level and two high-level (expressions and visimes) parameters specified for the facial animations as well as 296 parameters specified for the body animation. In this way, ECA developers have to calculate inverse kinematics to animate the character. Some software packages are available for MPEG-4 FBAP, for example—visage|SDK (http://www.visagetechnologies.com).

SAIBA Framework

To scaffold the ECA production process and encourage sharing and collaboration, a group of ECA researchers has initiated the SAIBA framework (Situation, Agent, Intention, Behavior, Animation) (mindmakers.org, 2006). The framework specifies multimodal generation and consists of processing stages in three different levels.

1 Planning of a communicative intent.
2 Planning of a multimodal realization of this intent.
3 Realization of the planned behaviors.

This working group aims to provide two common languages for describing ECAs. One serves as the interface between stage 1 and 2. They call this Function Markup Language (FML). The other one is the interface between stage 2 and 3, which they call Behavior Markup Language (BML).

Function Markup Language

Function Markup Language is a language that describes the communicative and expressive intention of ECAs without any reference to physical behavior. It is meant to provide a semantic description that accounts for the aspects that are relevant and influential in the planning of verbal and nonverbal behaviors. The specification of FML is still at its beginning stage. The first FML workshop was held together with the AAMAS 2008 conference, which the author of this chapter also attended. In this workshop, the researchers discussed the range that FML should cover. It discussed issues such as what FML actually means, what the term intention means, and whether culture, emotion, personality, or context should be included as well. The discussion was started from a very abstract view and there was no concrete agreement achieved in this workshop, but the researchers agreed to form smaller groups to develop proposals based on four specific scenarios. These scenarios include:

- Dyadic conversation with a human in a scenario where the agent is collaborating with the user on the construction of a physical object. Negotiations involving topics such as what to do in the next step to achieve the goal are expected.
- An agent presenting a science exhibit to visitors at a science museum. It is considered to be a “long” monologue, and the agent is assumed not be able to perceive the audience feedback.

- Multiparty conversation in social interactions that are expected to take place in a restaurant. Participants in the conversation are assumed to be dynamic, i.e. the participants may join and leave the conversation while it is continuing.
- A long-term companion agent in the health domain. The scenario will describe two to three interactions at widely separated points in time during this long-term relationship.

However, after the workshop, there was no further concrete progress in the development of FML. This is probably due to the range of concepts that FML has to cover being too broad. Researchers from various fields have large variety of the demands and requirements for FML. The lack of a strong and centralized organization also increased the difficulty in achieving concrete results.

Behavior Markup Language

Behavior Markup Language is a language meant to describe multimodal behaviors as they are realized in the final stage of the generation process. It provides a general, player-independent description of multimodal behavior that can be used to control an ECA. In contrast with FML, the aspects that BML is designed to address are much more concrete and clear. The working group first proposed the idea in Kopp, Bergmann, and Wachsmuch (2006) and discussed their progress and some specific technical issues in Vilhjálmsson et al. (2007). A draft specification (Mindmakers.org, 2008) has been published.

It may be distinguished from the other languages mentioned above in that it mainly proposes syntax describing the synchronization of multiple modalities of the character. In BML, a concept called "synch point" is proposed. Each individual nonverbal action of the character has a single ID and six phases, which are divided by five points: Start, Ready, Stroke-start, Stroke-end, Relax, and End. Speech texts are inserted with synchronization marks. The synchronization of multimodal animation is then described via the alignment of these synch points by referencing the action IDs. In BML, the working group defined the character animations as the following core categories: posture, locomotion, speech, gesture, face, head, and gaze. Each category has its own set of XML elements and attributes and has a minimum set of animations, which must be implemented by any BML-compliant player that the developers call the level 0 of BML. Work on BML is still in its progress and the specification is changing. Many parts are still missing or left unclear—for instance locomotion that has a target, such as walking and facial expressions. Currently, facial expressions seem to be specified with detailed parameters based on FACS (Facial Action Coding System) (Ekman & Friesen, 1978). This is an incoherent approach to nonverbal animation specifications in BML, which are merely specified with abstract terms like nodding and shaking of the head. Although the BML specification is not yet complete, several institutes have started works related to it. The virtual human team at the University of Southern California has developed a BML-compliant inverse-kinematics engine called SmartBody (Thiebaux, Marshall, Marsella, & Kallmann, 2008). The team in Reykjavik University developed a BML realizer by combining SmartBody and the free 3D graphics engine Panda3D (http://www.panda3d.org/) developed at Carnegie Mellon University.

Common Framework for the Development of ECAs

To achieve a believable ECA capable of natural face-to-face and multimodal conversation with humans is not easy. In addition to the prosodic properties of the verbal channel, precise control of nonverbal channels like gazing, raising of eyebrows, nod, hand gestures or postures in performing communication can be added. Functions can direct the flow of conversations or, as a supplement to verbal utterances, reflect the agent's internal emotional state, personality, and social status in response to recognized attention from human users with sensing devices. Finally, output can be rendered with realistic characters and fluent speech synthesis. To realize these abilities with a software agent, the knowledge and techniques of signal processing, natural language processing, gesture recognition, artificial intelligence, dialog management, personality, and emotion modeling, natural language generation, gesture generation, CG character animation and so on are required. Embodied conversational agents involve so many research disciplines that it is difficult for individual research teams to develop them from scratch. No matter what field developers who are going to build an ECA are in, they need to include a minimum set of these functionalities into their ECA. The usual way to build ECA systems is therefore by using software tools developed by other research groups. However, because software tools developed by different institutes are neither meant to cooperate with each other nor designed for the same application domain, it is usually laborious or even impossible to make them work together. More than that, similar approaches are repeated by researchers due to their common needs, leading to redundant work.

To solve these problems, a common framework is needed that absorbs the heterogeneities to connect diverse ECA software tools and drives the connected components as an integral ECA system. Then redundant efforts can be avoided and resource can be saved. Furthermore, the sharing of research results can be facilitated and the development of ECA systems can become easier.

Like the typical modeling of regular autonomous agents, an ECA needs to possess the following abilities:

- it should perceive verbal and nonverbal inputs from the user and the environment where the user is;
- it should interpret the meaning of the inputs and deliberate appropriate verbal and nonverbal actions as the responses;
- it should perform those actions with an animated computer graphic character in a virtual environment.

In order to realize these abilities, various functionalities like acquiring sensor data, speech recognition, gesture recognition, understanding natural language, BDI planning, speech synthesizing, CG character animation, and so on are required. Here, we call the modules that handle each individual function components of the whole ECA system. In a 2002 workshop (Gratch, Rickel, Andre, Cassell, Petajan, & Badler, 2002), around 30 international ECA researchers had intensive discussions about how to achieve modular architecture and interface standards that will allow researchers in this area to reuse each other's work. However, this goal has not yet been realized, except work on the SAIBA framework.

To achieve a common ECA component framework for general purposes, there are various requirements that should be fulfilled.

- **Integration platform:** a platform that can seamlessly integrate various ECA components and drive them to jointly behave as an integral ECA is indispensable.
- **Distributed and OS/programming language independence:** components may be developed in various programming languages and run on various operating systems. Hence, it is necessary for the integration framework to cover major operating systems and programing languages and allow the connected components to run on multiple machines.
- **Modularity and reusability:** this should be the heart of any integration approach. Component reusability can be maximized by clearly dividing the functionalities of components and clearly defining the interfaces between them. Simpler functionalities handled by each component and lower interdependency improve modularity.
- **Support for various natural languages:** with advances in transportation the world has become smaller and smaller, and crossculture issues have been emerging as much more important than ever before. However, due to the fact that Western countries dominate the development of the computer science field, issues related to Asian languages or others are often ignored. Flexibility to handle various languages needs to be maintained.
- **Two-way communication among components:** the ECA components do not only "pull" data from the others but some of them, such as sensor data processing components, also have to "push" data to the others. Hence a mechanism that supports two-way data passing is required.
- **Real-time performance and timing control:** real-time response of the agent to the user's inputs is one of the basic requirements of ECA systems. The latency of each part of the system needs to be kept at a minimum while on-time execution of actions need to be guaranteed. Therefore, a strict temporal model is a necessity.
- **Ease the efforts to adopt legacy systems:** libraries and tools should be provided to ease the efforts to develop wrappers for adopting legacy systems to be connected to the architecture.
- **Support for ECA-specific functionalities:** in contrast to general-purpose distributed architecture, for an architecture dedicated to the development of ECAs, the following supports are required.
- **Fusion of multimodal inputs:** in multimodal interactive ECA systems, the relationship of user inputs from the speech channel and other possible sensory channels needs to be identified correctly and trigger appropriate responses from the agent.
- **Synchronization between prerecorded tracks and run-time generated behaviors in outputs:** fixed length prerecorded tracks such as voice, music, or motion captured animation sequences need to be synchronized with variant length run-time generated animations.
- **Synchronization between verbal and nonverbal behaviors in outputs:** verbal and nonverbal behaviors are interrelated, supple each other and need to be synchronized.
- **Virtual environment control:** not only the virtual character itself but also the virtual environment that it lives in need to be altered corresponding to the

interactions between the agent and the human user, as in scene changes and camera manipulations.

- **User interruption:** provide the flexibility that allows a smarter system to modify its current behaviors online instead of simply stopping them and then launching new ones.

In order to meet these conditions, our group proposed the Generic Embodied Conversational Agent (GECA) development framework (Huang, Cerekovic, Nakano, Pandzic, & Nishida, 2008). It integrates distributed and reusable ECA modules so that they behave as an integral agent. It is composed of three parts. GECA Platform is a network communication middleware based on a blackboard and XML message exchanging. It provides services including a naming service, message subscription, and message forwarding management. GECA Plugs are the libraries that absorb the differences among operating systems and programming languages to facilitate the development of the wrappers of individual ECA components. GECA Protocol (GECAP) is a specification of XML message types and formats that are exchanged among the components. Based on this framework, GECA Scenario Markup Language (GSML), describing human-agent interactions, and its execution component were developed to supplement GECAP. GSML is an XML-based script language to define a state transition model for a multimodal dialog between the user and the agent.

Conclusions

As the media equation theory (Reeves & Nass, 1996) suggests, people tend to treat computers and other media as if they have personalities. Virtual agents can serve as the embodiment of the target. Although they do not necessarily have a graphical character, like Apple's Siri, Google's Google Now, and Microsoft's Cortana, more and more personal assistant agents have been being introduced as the interface between human users and information systems. The dialogue system simplifies the input procedure and the character animation enables the system's feedback to be understood more intuitively. Moreover, they do not only provide a shell that helps the user to use the system easier but also lead to attachment toward the systems. Around 20 years since the launch of this research field, many research issues have been explored. The output phase and input phase of ECAs saw tremendous achievements. Ultrarealistic character animation and speech synthesis with rich emotional expressiveness is already available. However, the core part of ECAs, that is, the functions of their *intelligence*, are still far from a satisfying level and leave a large space for exploration. We expect that research topics in this field will continue to attract researchers for years.

References

Arafa, Y., & Mamdani, A. (2003). Scripting embodied agents behaviour with cml: character markup language. In *Proceedings of the Eighth International Conference on Intelligent User Interfaces* (pp. 313–316). New York, NY: ACM.

Becker, C., Prendinger, H., Ishizuka, M., & Wachsmuth, I. (2005). Evaluating affective feedback of the 3d agent max in a competitive cards game. In *Proceedings of the First*

International Conference on Affective Computing and Intelligent Interaction (ACII'05) (pp. 466–473). Berlin: Springer.

Boukricha, H., Becker, C., & Wachsmuth, I. (2007, September). *Simulating empathy for the virtual human max.* Paper presented at the Second International Workshop on Emotion and Computing, in conjunction with the 30th German Conference on Artificial Intelligence (KI'07), Osnabruck, Germany.

Carolis, B. D., Pelachaud, C., Poggi, I., & Steedman, M. (2004). APML, a mark-up language for believable behavior generation. In H. Prendinger & M. Ishizuka (Eds.), *Life-like characters. tools, affective functions, and applications* (pp. 65–85). New York, NY: Springer.

Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjálmsson, H., & Yan, H. (1999, May). Embodiment in conversational interfaces: Rea. In *Proceedings of the 17th international conference on human factors in computing systems CHI'99* (pp. 520–527). New York, NY: ACM.

Cassell, J., Bickmore, T., Campbell, L., Vilhjálmsson, H., & Yan, H. (2000a). Human conversation as a system framework. In J. Cassell, J. Sullivan, S. Prevost, & E. Churchill (Eds.), *Embodied conversational agents* (pp. 29–63). Cambridge, MA: MIT Press.

Cassell, J., Stocky, T., Bickmore, T., Gao, Y., Nakano, Y., Ryokai, K.,… Vilhjálmsson, H. (2002, February). MACK: Media lab autonomous conversational kiosk. In *Proceedings of Imagina'02: Intelligent Autonomous Agents* (vol. 2, pp. 12–15). Cambridge, MA: MIT Media Lab.

Cassell, J., Sullivan, J., Prevost, S., & Churchill, E. (Eds.). (2000). *Embodied conversational agents.* Cambridge, MA: MIT Press.

DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A.,… Morency, L.-P. (2014). Simsensei kiosk: A virtual human interviewer for healthcare decision support. In A. Lomuscio, P. Scerri, A. Bazzan & M. Huhns (Eds.), *Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014)* (pp. 1061–1068). New York, NY: ACM.

Ekman, P., & Friesen, W. (1978). *Facial action coding system.* Palo Alto, CA: Consulting Psychologists Press.

Gaschler, A., Kessler, I., Petrick, R. P. A., & Knoll, A. (2015). Extending the knowledge of volumes approach to robot task planning with efficient geometric predicates. In *IEEE International Conference on Robotics and Automation (ICRA)* (pp. 3061–3066). doi:10.1109/ICRA.2015.7139619

Gratch, J., Rickel, J., Andre, E., Cassell, J., Petajan, E., & Badler, N. (2002). Creating interactive virtual humans: Some assembly required. *IEEE Intelligent Systems*, *17*(4), 54–63.

Gustavsson, C., Beard, S., Strindlund, L., Huynh, Q., Wiknertz, E., Marriott, A., & Stallo, J. (2001). *VHML (Working Draft v0.3 ed.)* [Computer software manual]. Retrieved from http://vhml.org/vhml_about/html/vhml_about.shtml

Hayashi, M., Ueda, H., Kurihara, T., & Yasumura, M. (1999, January). TVML (TV Program Making Language). Automatic TV program generation from text-based script. In *IEICE technical report* (Vol. 98, pp. 1–8). Tokyo, Japan: Institute of Electronics, Information and Communication Engineers.

Huang, H. H., Cerekovic, A., Nakano, Y., Pandzic, I. S., & Nishida, T. (2008). The design of a generic framework for integrating ECA components. In L. Padgham, D. Parkes, & J. P. Muller (Eds.). In *Seventh International Conference of Autonomous Agents and Multiagent Systems (AAMAS'08)* (pp. 128–135). Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

Huang, H. H., Furukawa, T., Ohashi, H., Cerekovic, A., Pandzic, I. S., Nakano, Y., & Nishida, T. (2010). How multiple current users react to a quiz agent attentive to the dynamics of their participation. In *Ninth International Conference on Autonomous Agents and Multiagent Systems (AAMAS'10)* (pp. 1281–1288). Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

Huang, H. H., Uejo, M., Seki, Y., Lee, J.-H., & Kawagoe, K. (2012). Measuring user performance for a virtual ballroom dance instructor to generate real-time feedbacks. *ICIC Express Letter*, *6*(12), 3077–3086.

Jones, H., Sabouret, N., Damian, I., Baur, T., Andrre, E., Porayska-Pomsta, K., & Rizzo, P. (2014, February 24). *Interpreting social cues to generate credible affective reactions of virtual job interviewers.* Paper presented at the Second International Workshop on Intelligent Digital Games for Empowerment and Inclusion, Haifa, Israel.

Kendon, A. (2004). *Gesture: Visible actions as utterance.* Cambridge: Cambridge University Press.

Kopp, S., Allwood, J., Grammer, K., Ahlsen, E., & Stocksmeier, T. (2008). Modeling communication with robots and virtual humans. In *Proceedings of the Second ZiF Research Group International Workshop on Embodied Communication in Humans and Machines* (pp. 18–37). Berlin: Springer.

Kopp, S., Bergmann, K., & Wachsmuch, I. (2008). Multimodal communication from multimodal thinking—towards an integrated model of speech and gesture production. *International Journal on Semantic Computing*, *2*(1), 115–136. doi:10.1142/S1793351X08000361

Kopp, S., Gesellensetter, L., Kramer, N. C., & Wachsmuth, I. (2005). A conversational agent as museum guide—design and evaluation of a real-world application. In *Proceedings of the Fifth International Conference on Intelligent Virtual Agents (IVA'05)*. Berlin: Springer.

Kopp, S., Jung, B., Lebmann, N., & Wachsmuth, I. (2003). Max—a multimodal assistant in virtual reality construction. *Künstliche Intelligenz*, *3*(4), 11–17.

Kopp, S., Krenn, B., Marsella, S., Marshall, A. N., Pelachaud, C., Pirker, H.,… Vilhjálmsson, H. (2006). Towards a common framework for multimodal generation: The behavior markup language. In *Proceedings of the Sixth International Conference on Intelligent Virtual Agents (IVA'06)* (pp. 205–217). Berlin: Springer.

Kranstedt, A., Kopp, S., & Wachsmuth, I. (2002). *Murml: A multimodal utterance representation markup language for conversational agents* (Report No. SFB 360 Situated Artificial Communicators). Bielefeld: Bielefeld University.

Kshirsagar, S., Magnenat-Thalmann, N., Guye-Vuilleme, A., Thalmann, D., Kamyab, K., & Mamdani, E. (2002, May). *Avatar markup language.* Paper presented at the Eighth Eurographics Workshop on Virtual Environments, Barcelona, Spain.

Mindmakers.org. (2006). *SAIBA multimodal behavior generation framework.* Retrieved from http://www.mindmakers.org/projects/saiba/wiki

Mindmakers.org. (2008, August). *Behavior markup language (bml) version 1.0 (draft).* Retrieved from http://www.mindmakers.org/projects/bml-1-0/wiki

Mori, M. (1970). The uncanny valley. *Energy*, *7*(4), 33–35.

Nakano, Y. I., Reinstein, G., Stocky, T., & Cassell, J. (2003). Towards a model of face to face grounding. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics (ACL'03)* (pp. 553–561). doi:10.3115/1075096.1075166

Nishida, T. (Ed.). (2007). *Conversational informatics: An engineering approach.* New York, NY: John Wiley & Sons, Inc.

Okazaki, N., Aya, S., Saeyor, S., & Ishizuka, M. (2002, November). *A multimodal presentation markup language MPML-VR for a 3d virtual space.* Paper presented at the Workshop on Virtual Conversational Characters: Applications, Methods, and Research Challenges, Melbourne, Australia.

Pandzic, I. S., & Forchheimer, R. (Eds.). (2002). *Mpeg-4 facial animation, the standard, implementation and applications.* Chichester: John Wiley & Sons, Ltd.

Pelachaud, C., Carofiglio, V., Rosis, F. de, & Poggi, I. (2002). Embodied contextual agent in information delivering application. In *Proceedings of the First International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS'02)* (pp. 758–765). doi: 10.1145/544862.544921

Prendinger, H., Descamps, S., & Ishizuka, M. (2002). Scripting affective communication with life-like characters in web-based interaction systems. *Applied Artificial Intelligence*, *16*(7–8), 519–553.

Prendinger, H., & Ishizuka, M. (Eds.). (2004). *Life-like characters—tools, affective functions, and applications.* Berlin: Springer.

Reeves, B., & Nass, C. (1996). *Media equation: How people treat computers, television and new media like real people and places.* Chicago, IL: University of Chicago Press.

Thiebaux, M., Marshall, A. N., Marsella, S., & Kallmann, M. (2008). Smart-body: Behavior realization for embodied conversational agents. In *Proceedings of the Seventh International Conference of Autonomous Agents and Multiagent Systems (AAMAS'08)* (pp. 151–158). Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems.

Traum, D., Bos, J., Cooper, R., Larsson, S., Lewin, I., Matheson, C., & Poesio, M. (1999). *A model of dialogue moves and information state revision* (Report No. D2.1). Gothenburg, Sweden: Department of linguistics, Goteborg University. Retrieved from Trindi Project website: http://www.ling.gu.se/projekt/trindi/private/deliverables/D2.1/D2.1.pdf

Vilhjálmsson, H., Cantelmo, N., Cassell, J., Chafai, N. E., Kipp, M., Kopp, S.,… van der Werf, R. J. (2007). The behavior markup language: Recent developments and challenges. In *Proceedings of the 7th international conference on intelligent virtual agents (IVA'07)* (pp. 99–111). Paris, France: Springer.

Web3D Consortium. (1997). *The virtual reality modeling language (VRML)* (ISO/IEC 14772-1:1997). Retrieved from https://www.iso.org/standard/25508.html

Zlotowski, J. A., Sumioka, H., Nishio, S., Glas, D. F., Bartneck, C., & Ishiguro, H. (2015). Persistence of the uncanny valley: The influence of repeated interactions and a robot's attitude on its perception. *Frontiers in Psychology*, *6*, 883.

27

Interacting with Mobile Media

S. Shyam Sundar, Eugene Cho, and Jinping Wang

Mobile devices have replaced desktop and laptop computers as the principal media for interaction and communication. According to a Pew report, nearly two-thirds of Americans own a smartphone and "for many, these devices are a key entry point to the online world" (Smith, 2015). Aside from making phone calls and exchanging text messages with others, smartphone users look up information on the Internet, listen to music, watch videos, play games, conduct financial transactions, take online classes, apply for jobs, and engage in a whole range of activities that were previously conducted on a computer.

With the widespread adoption and diffusion of smartphones and tablets, the landscape of HCI has changed dramatically, just in the last 10 years. Conferences and journals devoted exclusively to mobile interaction have emerged, and even mainstream HCI research has taken a significant turn toward mobile interfaces. Increasingly, design as well as user studies in HCI assume a mobile device.

Given the considerable difference in form factor, there has been a renewed interest in the basics of user interaction—methods of text entry, input and output modalities, and display size. In addition, mobile media have introduced new concepts to the HCI lexicon, such as mobility, and altered the definitions of existing concepts such as ubiquity, synchronicity, and propinquity. Research has also focused on the use of mobile media in such higher order domains as information-seeking, entertainment, social interactions, advertising, and marketing, shopping and commerce, health, human welfare, and economic development.

This chapter will first describe research pertaining to mobile interfaces, with a focus on input modalities and formal features such as screen size and interaction techniques. It will then discuss application domains of mobile HCI.

Mobile Interface Design and Usability

The starting point of all interaction is user input. As a field, HCI has a rich history of research and development of input modalities and interaction techniques. While keyboard-based input predominated in early years, mouse input proved to be a powerful

The Wiley Handbook of Human Computer Interaction, Volume 2, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

tool for interacting with graphical user interfaces (GUIs) by affording pointing, clicking, and dragging. Web-based media extended the range of interaction techniques to sliding, zooming, hovering, and flipping (Sundar, Bellur, Oh, Xu, & Jia, 2014). Much of this has been replicated with mobile interfaces, but with a special emphasis on haptic modality, because touch has emerged as the dominant method for initiating interaction with a mobile device.

Touch-based interaction

Studies in the HCI literature tend to focus on touch-based features, including text entry, and evaluate the efficiency of various touch-based techniques and analyze user behaviors, in order to inform design and development of mobile interfaces.

The limited screen size of mobile devices has inspired the design of efficient target acquisition techniques such as Linear Dragging (Au, Su, & Lau, 2014), which enables users to select items from dense visual spaces with the touch of a finger. While extensive work on mobile touch screens has tended to focus on smaller devices, the increasing prevalence of tablets and larger smartphones has triggered interest in touch-based interaction with larger mobile screens (Girouard et al., 2015; Wolf & Henze, 2014).

While larger screens afford more visual space for interaction, the form factor compromises ease of operation because of the difficulty of inputting while simultaneously holding the device, all with a single hand. Smaller mobile devices easily allow single-hand interaction, but larger screens limit users' ability to place their thumb on the target effectively, thus leading to the use of both hands or the device slipping from the user's hand (Girouard et al., 2015). Nonetheless, one-handed smartphone usage is the desired ideal. To achieve this, Girouard et al. (2015) examined how the different locations of the single-handed grip on smartphones (i.e., top left corner, top right corner, lower bottom, and center) and the direction of bend gestures (i.e., up, down) would affect the time of completion and user comfort associated with different touch-based tasks (e.g., answering a phone call, browsing photos), and found that the top right-up bend and the center-up bend gestures are the fastest and most preferred. Tablets with larger screens also led to other indirect touch strategies needed when gripping the device with "both" hands. For example, Wolf and Henze (2014) found that using a miniature interaction area (i.e., "miniature representation of the entire screen"; p. 56) for touching the target could serve as an effective alternative.

Aside from simple pointing techniques, several hand gestures on multitouch screens have been designed and tested over the years. Based on Web-browsing behaviors, Park and Han (2014) proposed an analytical approach to the creation of multitouch gesture vocabularies applicable to mobile devices, using a combination of tapping, pressing, flicking, dragging, and drawing. Poppinga, Shirazi, Henze, Heuten, and Boll (2014) found that the "most frequently used gesture activator is the lock screen, followed by the notification bar, the activity, and the wallpaper" (p. 180). In general, letter-shaped gestures are preferred in touch-based gestures.

Tu, Ren, Tian, and Wang (2014) tested two specific scrolling techniques, flick (e.g., linear scrolling on iPhones) and ring (i.e., circular scrolling for iPods), and discovered that flick is generally more efficient (especially in terms of lowering the number of crossings where the user enters or leaves a specified area) than ring for both sitting and walking postures across three input methods (i.e., index finger, pen, and thumb) during document-navigation tasks. Dou and Sundar (2016) found that adding a swiping interaction technique to a tap-only mobile website enhanced user experience by heightening perceived enjoyment and self-reported behavioral intentions to return to the website.

Other studies have examined the efficiency of touch-based features applied to a specific smartphone user population (e.g. elderly users; Hwangbo, Yoon, Jin, Han, & Ji, 2013). For example, when elderly users were instructed to touch a square target shown on screen surrounded by other squares, Hwangbo et al. (2013) found that, as the target size grew larger and the spacing between squares became wider, the task completion time and the error rate decreased. More important, a significant interaction between target size and spacing appeared, and revealed that when the target size was 12 mm (vs. 5 mm and 8 mm), the smaller spacing of 1 mm (vs. 3 mm) led to decreased task completion time and error rate, which suggests a layout for elderly users with narrow target spacing for a certain level of target size. In addition, pointing performances were improved by audiotactile feedback (vs. no feedback or tactile feedback) in terms of both task completion time and number of errors.

Touch interfaces have also been developed for a variety of specific purposes, ranging from language learning to biometric authentication. For example, Veras et al. (2014) introduced a multitouch tablet interface specifically designed to aid English vocabulary acquisition with particular functions, such as highlight features on unknown words (with fading features on known words), synonym replacement and ranking features, and word definition offerings. Touch-based features are also applied to biometric authentication using a fingerprint scanner (e.g., Apple's Touch ID and Android's Face Unlock), with users overlooking privacy concerns in favor of better usability (De Luca, Hang, von Zezschwitz, & Hussmann, 2015).

Text entry interfaces

Another aspect of touch-based interaction that has received considerable attention is text entry. Virtual keyboards have been the focus of several studies. As a recent example, Henze, Rukzio, and Boll (2012) attempted to expand previous studies on users' performance that were concentrated on certain keyboard functions (e.g., key size, keyboard layout). In particular, the researchers collected keystroke data from a typing game based on an Android keyboard system, and according to the general skew on keyboard touch distribution toward the bottom of the screen, found that shifting the users' touch events toward the upper part of the screen could enhance user performance for existing Android keyboards in terms of speed and error rate at no additional cost.

Romano, Paolino, Tortora, and Vitiello (2014) suggested a new interface that offers advantages over traditional QWERTY keyboards for entering text in a limited screen space. Named Tap-and-slide (TaS), it allows the virtual keyboard to work with only six buttons (each with a cluster of letters) with two gestures (i.e., tap and slide) and was found to be not only easy to learn but also more accurate than traditional QWERTY keyboards.

Interestingly, several studies about mobile input pertain to Chinese characters, due to the language's distinctive characteristics. Recent findings suggest that handwriting input method results in shorter task completion, higher perceived ease of use and intention to use than typing into an onscreen keyboard, especially when the mobile screen is small (Zhou, Rau, & Salvendy; 2014), with scholars identifying an optimal size of input box (i.e., 2.5 × 2.5 cm) for one- and two-handed Chinese handwriting input (Tu & Ren, 2013). Noting that current keyboards are designed for the Latin alphabet, Niu, Liu, Lin, Zhu, and Wang (2014) suggested an alternative keyboard layout (named Stroke+ +) based on the hieroglyphic properties of Chinese characters, which outperformed handwriting input in terms of both efficiency and usability.

With SMS, instant messaging, social media and numerous other applications that require text input, there exists a great need for improving the ease and efficiency of text entry into mobile devices. This is likely to remain an active area for design innovation and usability evaluation in the future.

Speech recognition

Aside from touch and text, speech is likely to emerge as the most dominant input modality for mobile interfaces, considering that users are naturally inclined to speak into their smartphones. Interface agents such as Siri invite speech input and aim for a seamless interaction. Speech has several advantages as an input modality, such as freeing users from typing and paying visual attention to the device. Shirali-Shahreza, Penn, Balakrishnan, and Ganjali (2013) compared speech to soft keyboards for entering short text and they found out that when users see the word and say it to the phone (SeeSay CAPTCHA), it "requires less time to be solved and users prefer it over current text-based CAPTCHA methods" (p. 2147).

Gaze as input modality

Gaze is a new modality for small-screen handheld devices, but appears to require a different cadence of operation by the user, leading to usability challenges. Rozado, Moreno, San Agustin, Rodriguez, and Varona (2015) suggested and evaluated a video-based eye-tracking system that can interact with handheld devices. It was found that when users were instructed to engage in a series of gaze sequence gestures (e.g., move gaze from leftup to rightup to rightdown to leftdown part of the smartphone screen), those who started and ended with a dwell (vs. no dwell) while performing their gaze gesture were slower and perceived it as tiring, but were more accurate in performing their tasks.

Gestures and motion as input

Finally, motion is emerging as an input modality. There appears to be a broad consensus on the motion gestures preferred by users for many common actions, such as answering the phone, ignoring a call, or navigating within applications, leading researchers to develop a taxonomy for motion gestures like clutching and shaking the phone, bringing it closer to face or mouth, and rotating the device along different axes (Ruiz, Li, & Lank, 2011). Another application of motion as an input modality is in using the smartphone as a remote control device to experience public displays. For example, Bergé, Serrano, Perelman and Dubois (2014) demonstrated that mid-air gestures with smartphones, i.e., hand motions around the phone to obtain overview and detail from 3D public displays, "are more efficient and preferred than touchscreen input" (p. 132).

In sum, several lines of HCI research are investigating various innovative input modalities for interacting with and via mobile interfaces. However, many of them remain in the prototype stage and are yet to be fully incorporated into commercially available mobile devices. Instead, the trends that are more prevalent in the marketplace have to do with screen size, graphics and context awareness, which we describe in the sections that follow.

Display size and type

The size of display has been a major concern with manufacturers of smartphones and tablets. Screen sizes have increased in recent years, providing users with more real estate to conduct transactions and enjoy high-fidelity pictures, graphics, and videos. While large screens have been associated with greater enjoyment, they are also perceived as being lower in mobility (Kim, Sundar, & Park, 2011). Studies have shown that interacting with larger screens could lead to increased efficiency when seeking information (Raptis, Tselios, Kjeldskov, & Skov, 2013), positive attitudes (mediated by perceived affective quality) and heightened perceived usefulness (mediated by perceived control and ease of use), which ultimately increased intention to use the mobile device (Kim & Sundar, 2014). Larger screens are known to encourage heuristic over systematic processing, resulting in greater affective and behavioral trust in the content, with consequences for user reception of persuasive content such as mobile ads (Kim & Sundar, 2016). The type of display panels (i.e., IPS vs. AMOLED), however, appears not to have an effect on smartphone users' perceived levels of "oculomotor comfort, enjoyment, display quality, viewing satisfaction, presence, text readability, reading comprehension, and reading satisfaction" (p. 77), despite claims to the contrary by mobile device manufacturers (Kim, Park, & Sundar, 2012).

Graphics and 3D/VR/AR features

With larger screen sizes and greater fidelity in rendering graphics, images, and video content, mobile media have emerged as a viable platform for 3D and other immersive-reality experiences. In mobile interfaces, 3D is not simply a display technology but one that can serve as an interactivity tool. Häkkilä et al. (2014) tried to enable cinema audience members to retrieve content from a 3D screen by using their mobile devices during movie viewing, so that they could view additional details about onscreen activities and related promotional opportunities on their devices.

Caballero, Chang, Menéndez, and Occhialini (2010) proposed an augmented virtuality prototype called "Behand," which allows users to manipulate virtual 3D objects within the phone with hand gestures to carry out a number of activities, ranging from gaming to sketching to collaboration. Li, Zhang, Sundar, and Duh (2013) developed an augmented reality game for the tablet by providing users an overlay of historical information about certain sights in the city of Singapore. This also included nonplaying characters, ostensibly original immigrants, who helped provide clues to users. A controlled experiment revealed that such location-based AR can indeed improve student learning by engaging them deeply with lesson content.

Since the mid-2000s there has been a number of AR apps for mobile devices, with millions of downloads, but it was Pokémon Go in the summer of 2016 that put mobile AR on the map, with the public bemoaning the side effects of its use in public places, echoing some of the findings by Hofmann and Mosemghvdlishvili (2014), showing that "AR users actually became less aware of their surroundings than non-AR navigational app users" (p. 265). On the positive side, mobile AR was shown by them to "improve users' understanding of proximity, directions and spatial relations" (p. 266).

Location-awareness

Location-awareness technology has also come of age, with profound implications for spatiality and sociality (de Souza e Silva, 2013). Some of the most popular apps on smartphones rely on user location in order to provide services. Location-based social media, such as Foursquare, allow users to "check in" at physical locations and connect with others as well as interact with a place (Kim & Lingel, 2016), although privacy remains a concern.

Location has become an important piece of information that people share with each other. Mobile geotagging is now a default choice for individuals posting pictures online. It offers a new way to build social familiarity with a space, to create narratives related to place, manage identities (Humphreys & Liao, 2011), and integrate the digital world with the material world (Hjorth & Pink, 2014). It is also known to serve an impression-formation function. Fitzpatrick, Birnholtz, and Gergle (2016) suggest that "impressions based on visual representations of the spatial self can be formed quickly and based on relatively few visible locations" (p. 301). An experiment that presented three check-in location markers with a photo of a stranger led to different impressions.

Apart from geotagging, location-based services (LBS) have expanded to many other fields in the form of specialized apps, each attracting its own share of research. A study of LBS retail apps, for example, suggests that higher perceived interactivity and compatibility promote consumers' affective involvement with the apps, leading to greater intention to download and use them (Kang, Mun, & Johnson, 2015). A study of news apps showed that they only scratch the surface of potential spatial connections to information by often restricting themselves to traffic and weather (Weiss, 2013). Designers have proposed plans for museums to adopt a "map-based approach that visualizes the location of exhibits" (p. 1070), which adjusts to the indoor location of visitors, while also providing access to more digital content pertaining to each of the exhibits (Wacker, Kreutz, Heller, & Borchers, 2016). Mobile dating and hookup apps based on location-awareness are proliferating. Birnholtz, Fitzpatrick, Handel, and Brubaker (2014) observed profile data in a location-aware social application for homosexual persons and found that users used precise language to avoid being connected to stigmatized behaviors yet conveyed signals for casual sex at the same time, and they were willing to share their location information without worrying about their privacy.

Location has therefore become an integral part of mobile media, connecting users to physical spaces—an especially important function given the mobility afforded by mobile devices. Just like screen size and immersive reality, location-based features serve to expand the boundaries of mobile HCI by providing users innovative affordances that inspire new apps and enrich their interactions with mobile media.

Mobile Application Domains

The application of mobile HCI has now expanded to almost all walks of life, from information-seeking to social interaction to health and human development, as evidenced by the burgeoning number of mobile apps in a number of domains. We describe some of the major application domains in the following sections.

Information-seeking

Information-seeking is perhaps the most important gratification sought and obtained from mobile media. It appears that most young people get their news via their mobile devices. Perceived relative advantage, ease of use, and usefulness have been identified as major predictors of news consumption among college students (Chan-Olmstead, Rim, & Zerba, 2013). The level of general online news consumption, preference for sports and technology news, and the level of mobile usage itself all drive mobile news usage (Chan-Olmstead et al., 2013). Van Damme, Courtois, Verbrugge, and De Marez (2015) found that majority of the mobile news originated from trusted major news brands, and mobile news seemed to offer users easy access to service-based news (e.g. weather, transportation) in the morning and when on the road.

Informational use of mobile media is also linked to civic participation, according to studies by Campbell and Kwak (2010, 2011a, 2011b). Using mobile phones for information exchange positively predicts the level of both civic engagement and political participation, albeit only among those who report higher levels of perceived competence with mobile telephony (Campbell & Kwak, 2010). In addition, using mobile phones for news is associated with public conversations with strangers (Campbell & Kwak, 2011a), and political discourse with close ties using mobile devices is associated with higher levels of political participation among those with large personal networks (Campbell & Kwak, 2011b).

Information-seeking uses of mobile media is common in educational settings and is found to have a positive impact on the learning environment. Access and dissemination of educational materials have been dramatically transformed due to mobile media (Shuib, Shamshirband, & Ismail, 2015). Data also show that students adopt note-taking apps on mobile devices for a wide range of functions, particularly “gathering and managing information, organization and planning, and the recording of ideas,” with the majority of students liking the software and finding it “quick and easy to use” (Schepman, Rodway, Beattie, & Lambert, 2012, p. 308). In addition to educational software, designers have devised new ways of enhancing learning and reducing anxiety by effectively leveraging touch-based interface for such activities as “visualizing unknown words” and “substituting difficult words with familiar synonyms” in a passage of text (Veras et al., 2014).

Given the social significance of disseminating information via mobile media, scholars have called for several design improvements, including “a standardized file format that is compatible with various platforms, better display technologies that are needed to enhance smartphone readability, and more stable and secure services” (Shin, Shin, Choo, & Beom, 2011, p. 2213). From a “digital divide” perspective, providing access alone is insufficient; “greater attention needs to be paid to designing devices that will enable underprivileged users to utilize a wider range of mobile communication applications” (Wijetunga, 2014, p. 722).

Entertainment

With increased affordance of video, mobile devices have become a major source of entertainment content, ranging from short YouTube video clips to live sports. Scholars have pointed out however that mobile video consumption is more than just watching TV on the go; it provides users with particular experiences, like sharing content in social situations, to create meaning and value (O'Hara, Mitchell, & Vorbau, 2007). In fact, social norms surrounding the use of mobile video appear to play a role in its adoption and use (Lin, Younbo, & Sim, 2015).

Designers have been working on developing better experiences and applications of mobile videos. For example, a mobile application called "MobileVideoTiles" enables groups of smartphone users to watch videos as a group on a big virtual screen composed of multiple mobile displays, thus enhancing the socialness of the shared viewing experience (Li, Scharf, & Kobbelt, 2016).

Users not only watch videos on mobile devices but also produce them. A group of teenagers tested a new prototype technology for live mobile video editing, while researchers observed how participants begin to master editing skills and focus on story creation. They argued that this kind of capacity to document and edit should be regarded as mobile video literacy, which emerged as it was introduced and adopted by users (Weilenmann, Säljö, & Engström, 2014)

Aside from video, games are a major source of entertainment on mobile devices. From Angry Birds to Pokémon Go, mobile games have enjoyed vast acceptance and use by the public. Affordances of mobile devices, such as interactivity cues that prime the presence of other players and haptic feedback for the user's moves, can positively affect user experience, especially game enjoyment (Lee, Bang, & Sundar, 2014). Spatial contexts of mobile device use, such as mobile context and home scenario, can also have an impact on the immersion and experience of players: those who play in a mobile scenario reported stronger immersion during game play, yet rated their experience more negatively than a home setting (Engl & Nacke, 2013). From a long-term perspective, good interaction in mobile games has a positive effect on a player's flow experience, leading to increasing player loyalty to the game applications (Su, Chiang, Lee, & Chang, 2016).

To improve user experience and game performance, designers have tested different input and output methods in mobile gaming. Touchless input like tilting serves to improve game performance, when compared with face tracking technology, which is perceived as more challenging and innovative by players (Cuaresma & MacKenzie, 2014). Cairns, Li, Wang, and Nordin (2014) argued that "slip control" led to a higher level of game immersion than either tilting or touching, because it gave users more sense of control. In terms of the output, some scholars focus on the effects of vibrations in mobile gaming. Esposito and Lenay (2011) introduced a FeelTact device, which was a type of tactile bracelet that could convey directions to players by tactile languages (e.g., two vibrations on the left bracelet being a signal of turning left). Players reported higher pleasure and immersion with such tactile feedback in an urban navigation game. Choe and Schumacher (2015) also found that higher intensity of vibration from the mobile device itself promoted perceived usefulness, and cognitive concentration of players in the game.

Games are also known to influence offline actions by users. For example, Frith (2013) documented how the gaming elements of location-based mobile games

(LBMGs) like Foursquare, such as securing badges, affect user decisions to visit certain locations. Jarusriboonchai, Malapaschas, and Olsson (2016) developed Who's Next, a mobile game intended to facilitate icebreaking activity among strangers, which turned out to be a success because it offered a relaxing way of sharing information about oneself and getting to know-strangers.

Inspired by the widespread appeal of games, academic research and design has tended to emphasize educational functions of mobile games (Furió, GonzáLez-Gancedo, Juan, Seguí, & Rando, 2013). One example is AREEF, a multiplayer underwater augmented reality (UWAR) experience designed to help kids gain a direct impression of fish and perform activities like catching "fish" and cleaning trash (Oppermann, Blum, & Shekow, 2016). Aside from augmented activities, stories in mobile games are found to be important for predicting the effectiveness of educational games and game satisfaction, especially for male students (Lu, Chang, Huang, & Chen, 2015). Aside from being teaching tools, mobile games can serve as testing tools to assess the cognitive abilities of players (e.g., Tong et al., 2015).

In sum, mobile media use for entertainment is dominated by video and games, providing a clear directive for designers to focus on these spaces, as they develop new tools and apps that result in a wide range of outcomes, from hedonic enjoyment to educational learning.

Mobile advertising and marketing

Aside from information and entertainment, other mass communication functions like persuasion are also being fulfilled by mobile media. Given the personal nature of mobile media, a particular challenge for marketers and advertisers is to go a step beyond vying for user attention and gain user trust, so that they can persuade individuals without appearing to be intrusive or distracting.

A major trend is personalization of marketing appeals. Tang, Liao, and Sun (2013) suggested a three-stage algorithmic framework for mobile personalized marketing (MPM) involving rule learning, rule selecting, and rule matching based on mobile users' contextual and interaction data. Similarly, Li and Du (2012) developed a targeted mobile advertising system (TMAS), a platform that links marketers and mobile device users based on personalization techniques. The system has three modules that interact with databases and users: (a) the advertisement management module, which allows marketers to revise content and properties of the ads, (b) user profile management module, which manages and updates user profiles modeled to predict interests and preferences of the consumers, and (c) an intelligent searching module, which personalizes search results based on user context (e.g., location, demographics, and preferences). Both these systems were tested and validated with real-time data.

Personalization based on user location is especially facilitated by mobile devices, leading to a plethora of applications intended to lure users to specific stores. Location-based advertising (LBA) is especially effective when there is buyin by the user in terms of tailoring the appeals and feature just-in-time information about shops in the vicinity selling products that are of interest to the user (Lee, Kim & Sundar, 2015). Such personalization helps reduce the perceived intrusiveness of LBA, which in turn positively influences attitudes toward LBAs in general and the ad in particular, as well as the intention to visit the advertised store.

Given the pronounced difficulty in garnering user interest in the mobile context, advertisers have had to be more creative and appeal directly to user emotions. Park and Salvendy (2012) found that mobile ads that elicit favorable emotions were related to positive attitude to ads and brands, and purchase intent, while stressful and apathetic ads linked to negative attitude towards ads and brand. Goh, Chu, and Wu (2015) found that, among people who were exposed to automobile ad campaigns in a mobile site, exposure to informative ad content (i.e., content heavily reliant on information) was negatively associated with depth of search, while positively promoting breadth of search. On the other hand, persuasive ads (i.e., content visually stimulating and hedonic) led to greater depth in search. Both the depth and breadth of search, in turn, positively influenced responses to ads (i.e., registering for a test drive).

Marketers have also examined the factors that make mobile ads go viral. In their study of mobile viral marketing campaigns (which encourage people to send and forward mobile text messages), and the possible factors that impact people's referral behaviors through mobile devices, Pescher, Reichhart and Spann (2014) found that consumers who place importance on entertaining and informing friends are more likely to be interested in and actually decide to engage in referring behaviors, while strength of tie is negatively associated with referent actions.

Beyond explicit promotional content on mobile devices, branded mobile applications has received research attention. For instance, Bellman, Potter, Treleaven-Hassard, Robinson, and Varan (2011) found that the use of branded apps improved consumers' attitude and purchase intentions toward the brands, with informational apps (i.e., Best Buy, Kraft, Target, Weber), being more effective in enhancing purchase intentions than experiential apps (i.e., BMW, Gap, Gillette, Lancôme). Kim, Wang, and Malthouse (2015) found that the adoption and continued use of a specific branded application, developed by a Canadian coalition loyalty program—Air Miles Reward Program (AMRP)—predicted increased spending levels (i.e., point accruals). Interestingly, two specific interactive features (i.e., information lookups and check ins) also led to increased spending.

In sum, advertising and marketing is a major application domain for mobile HCI posing significant challenges to designers, in terms of overcoming resistance and rejection by users. Affordances of personalization, interactivity, and context awareness hold significant promise in promoting positive user attitudes. Interestingly, both emotional and informational ads have been shown to be effective in the mobile realm, albeit via different theoretical mechanisms.

m-commerce

Mobile commerce is an application domain that continues to expand as more and more consumers as well as vendors conduct commercial transactions via mobile interfaces. Two streams of research are common in this area—m-commerce adoption (Chong, Chan, & Ooi, 2012; Gao, Waechter, & Bai, 2015; Okazaki & Mendez, 2013; Zhang, Zhu, & Liu, 2012) and mobile payment (Kim, Mirusmonov, & Lee, 2010; Liébana-Cabanillas, Sánchez-Fernández, & Muñoz-Leiva, 2014; Yang, Lu, Gupta, Cao, & Zhang, 2012).

Perceived security (i.e., trust) is a primary determinant of m-commerce adoption, with a variety of available services and social influence from friends, family, and mass media positively predicting consumer intentions (Chong et al., 2012). Portability and appealing interface design of mobile services predict ease of use, which in turn is associated with positive evaluations of mobile features (such as simultaneity, speed, and searchability), and perceived convenience of m-commerce (Okazaki & Mendez, 2013). Trust toward mobile vendors, the experience of flow state during mobile purchase activities, and satisfaction positively influence intentions to continue purchasing on mobile sites (Gao et al., 2015). In their meta-analysis of 53 articles, Zhang et al. (2012) found that perceived usefulness brought higher attitudinal change to use m-commerce in Western culture whereas perceived ease of use is more influential to increase perceived usefulness and behavioral intentions in Eastern culture for predicting attitudes toward m-commerce.

Perceived ease of use and perceived usefulness are also prominent in adoption studies of mobile payment. They are determined by several factors, such as innovativeness, m-payment knowledge, mobility, reachability, and convenience (Kim et al., 2010). Social influence, personal innovativeness, compatibility of mobile payment with personal lifestyle and relative advantage are all associated with higher intention to use m-payment (Yang et al., 2012). Liébana-Cabanillas et al. (2014) also found that the positive relationship between perceived trust and ease of use, and that between perceived trust and favorable attitude toward m-payment use were more pronounced among younger users, whereas social influence heightened the ease of use among older users.

Aside from mobile purchase and payment, m-commerce also includes customer service. Park and Sundar (2015) found that text-based interaction with a customer-service agent via instant messaging (IM) is a powerful tool in this regard. In their experiment, the provision of synchronous interaction served to enhance the sense of co-presence felt by the user with the agent whereas the use of emoticons gave users the impression that the agent understood their feelings. Both these outcomes in turn were positively associated with users' impressions of the task as well as the customer-service agent.

Mobile devices can also be used in an instrumental manner during the process of shopping by helping users create digital shopping lists. Heinrichs, Schreiber, and Schöning (2011) analyzed how people interact with paper-based shopping lists and designed elements related to collaborative creation and editing with family members, household vocabulary, and hybrid paper-digital medium into a mobile app called Digital Grocery List (DGL).

Mobile devices are being used for a number of m-commerce activities, from generating shopping lists to online shopping to making payments to interacting with customer-service agents. Nascent research points to the importance of trust and security, ease of use and social presence in transactions. Affordances that address these outcomes are likely to advance design as well as research in m-commerce.

Social interaction

Mobility and ubiquity of mobile devices has enabled people to stay more connected than ever before with their family, friends, and acquaintances. Texting is clearly the most preferred medium across all age groups (Forgays, Hyman, & Schreiber, 2014) and used for a variety of social interactions, including "microcoordination" (Ling & Lai, 2016). According to Gonzales (2014), text-based communication is associated with more self-disclosure than even face-to-face meetings or phone conversations, making it an intimate medium for social interaction. In general, users of short messaging services communicate more and feel more connected with each other (Lam, 2013). They turn to these services because they believe them to be more reliable and private (Church & de Oliveira, 2013), preferring messaging applications that do not have chat history (Rost et al., 2016). However, they consider instant messaging applications like WhatsApp to be more cost-effective than SMS (Church & de Oliveira, 2013). They also perceive a greater sense of immediacy and sense of community in WhatsApp compared to SMS.

Social interaction via mobile media is facilitated by emotional cues in the form of emoticons and emojis. These serve important functions in mobile communication—"adding emotional or situational meaning, adjusting tone, making a message more engaging, conversation management, and relationship maintenance" (Cramer, de Juan, & Tetreault, 2016, p. 504). Females use emoticons more frequently, while males use "a more diverse range of emoticons" (Tossell et al., 2012, p. 659). Design and research continue to expand the "emotional bandwidth" of mobile media. For example, Kang & Watt (2013) argued that increased anthropomorphism of avatars would result in increases in social co-presence and satisfaction with communication, without affecting the perceived social usefulness of mobile technology. Cui, Kangas, Holm, and Grassel (2013) designed a prototype to investigate the value of front-camera video recordings as emotional responses to photos shared within a group. It turned out that this method of interaction has positive effects on social interactions especially in geo-dispersed groups of young people. More generally, mobile communications involving video has become quite common for fostering stronger emotional connection, especially among couples in long-distance relationships (Greenberg & Neustaedter, 2013).

Perhaps the dominant platform for social interaction on mobile devices is mobile social media, which can be defined as "software, applications, or services accessed through mobile devices that allow users to connect with other people and to share information, news, and content" (Humphreys, 2013, p. 21). Recent years have seen a rise in the number of mobile apps for social sharing, especially those using the phone camera, with Snapchat and Instagram being the market leaders. Even traditional social media like Facebook are most often accessed via mobile devices (Sterling, 2016). Research suggests that accessing social networking sites via mobile devices is associated with feelings of being constantly connected to others and a higher sense of belonging (Quinn & Oldmeadow, 2013).

However, while mobile devices afford rich social interactions with distant others, they tend to diminish users' interactions with their proximate surroundings. An experiment by Banjo, Hu, and Sundar (2008) showed that individuals on a cell phone in a public place were less likely to offer help to people around them, i.e., proximate others. Other studies have raised safety concerns arising from users' tendency to multitask. One study showed that the habitual use of texting could be harmful since it would predict sending as well as reading texts on the road (Bayer & Campbell, 2012). Procyk, Neustaedter, Pang, Tang, and Judge (2014) found that it is difficult for mobile users to keep an eye on their surroundings during video chats, with potential to cause harm. Resolving this tension between mobile social interaction and one's proximate environment is an essential mandate for the design community.

m-health

A large body of mobile studies has focused on mobile phone overuse, addiction and dependency. Findings indicate that addictive and excessive mobile phone usage is associated with "cognitive failures in daily life" (Hadlington, 2015), reduced academic performance and wellbeing (Lepp, Barkley, & Karpinski, 2014; Li, Lepp, & Barkley, 2015), and anxiety when mobile devices are taken away (Cheever, Rosen, Carrier, & Chavez, 2014). Some of the major predictors of mobile addiction and problematic usage are external locus of control (Li et al., 2015), social self-efficacy, family, and emotional stress (Chiu, 2014), and simply the use of SNS mobile applications (Salehan & Negahban, 2013).

Another example of problematic mobile use pertains to sexting (Strassberg, Rullo, & Mackaronis, 2014), with 19.1% of high school students having sent and 38.2% having received a sexually explicit photo. Sexting among college students could mediate problematic alcohol use to result in sexual hookups among college students (Dir, Cyders, & Coskunpinar, 2013). In romantic relationships, Roberts and David (2016) found that the level of distraction caused by mobile phone usage when an individual is accompanied by his/her romantic partner (i.e., Pphubbing) is associated with relationship conflict, particularly among individuals with high (versus low) anxious attachment styles. This conflict is in turn associated with depression, which is mediated by relationship and life satisfaction.

Aside from problematic mobile usage, studies have focused on the opportunities and challenges related to mobile "fitness" devices and applications, with particular emphasis on the perceived quality of the data provided by them. El-Amrawy and Nounou (2015) attempted to objectively evaluate the performances of existing fitness devices and found that Misfit Shine showed the highest performance of reporting heart rates in terms of accuracy and precision among 17 wearable devices, while Samsung Gear showed the lowest in accuracy and Jawbone UP the lowest in precision. In a more user-focused approach, Yang, Shin, Newman, and Ackerman (2015) investigated how consumers evaluated their fitness tracking devices based on 600 product reviews on Amazon.com and 24 interviews, and found that precision was more important than accuracy for many users in reviewing the trend or pattern of their health data, while in some cases accuracy mattered for exercise optimization or medical reasons. In addition, Yumak and Pu (2013) identified 11 sensor-based health-tracking devices, which offered information on all four of the wellness dimensions (i.e., nutrition, exercise, sleep, and stress), and compared their value in terms of perceived usefulness (i.e., service coverage, prediction accuracy, abstraction and reflection), perceived ease-of-use (i.e., ease of device use, ease of software use, comfort), and perceived desirability (i.e., social acceptance, privacy, cost), with BeWell scoring the highest on overall performance.

Mobile apps addressing a variety of health issues have been developed. For example, Göllner, Hurtienne, Gollner, and Naumann (2011) came up with two design concepts—shared diary and collaborative meeting tool—to help deploy mobile technology for maintaining the independence and social activity level of people with dementia and older adults. Designers have also developed and evaluated gamified mobile health-related applications. For example, Chen and Pu (2014), based on their mobile game HealthyTogether, found that the social interaction through such an application promotes physical activities compared to exercising alone, and that cooperation-driven (versus competition) settings enhances physical activities, while Rooksby, Rost, Morrison, and Chalmers (2015) tested their mobile application, Pass The Ball (which tracks group members' physical activity based on a turn-taking mechanism, and lets them compete with other groups), and suggested implications for designing activity trackers: that is, attend to how people communicate about their own and others' data, and look beyond and try to understand the nuanced aspect of collaborative vs. competitive aspect of the design.

Applying mobile media in a more clinical setting, Chittaro and Sioni (2014) suggested that a certain modality (i.e., visualization over audio) embedded in a certain health-related application (i.e., breathing training app) could deliver better results (i.e., deeper breathing, increased user preference and perceived effectiveness). Focusing on aiding health workers, Dell, Francis, Sheppard, Simbi, and Borriello (2014) suggested a mobile camera-assisted system that runs affordable and prompt diagnostic tests in resource-constrained environments, and showed how such mobile systems could be integrated into clinical work flows and aid health workers in similar situations. Varshney (2014) suggested that future studies should address research challenges in terms of four categories (i.e., patients, healthcare professionals, information technologies, and applications) to reap more benefits from technological advances that m-health services offer. For example, for patients and healthcare workers, identifying efficient ways to deliver the right m-health service to the right patient as well as to promote healthy interaction between patients and professionals would be crucial. On the other hand, examining the technical issues related with advanced context-aware health applications and IT infrastructure is also necessary for future m-health research (Varshney, 2014).

The pervasiveness and ubiquitous nature of mobile phones seem to benefit those who take advantage of advanced mobile features to be more vigilant about their physical health and thus lead healthier lives. Future m-health studies are expected to widen our understanding of how we could use mobile technology to not only promote personal and public health but also enhance the quality of delivering medical care to patients.

Mobiles for development (M4D)

An extension of the use of mobile technology for health and wellbeing is an application domain known as Mobiles for Development (M4D), which is an outgrowth of two older streams of research–ICT4D (Information and Communication Technology for Development) and HCI4D (Human-Computer Interaction for Development). M4D is devoted to creatively leveraging the affordances of mobile technology for economic development in poor regions of the world and improving the lives of people of low socio-economic status.

Acknowledging M4D as a distinctive field of study (from ICT4D), but also an area in its nascent stage, Karippacheril, Nikayin, de Reuver, and Bouwman (2013) conducted interviews with 31 M4D experts to understand the characteristics of service platforms applied to M4D that connects service providers to end users. Three major mobile service platforms (MSPs) were identified (i.e., operators, devices, and service providers) to mediate services to the poor (i.e., "people living at the base of the economic pyramid (BOP)"; p. 24). Based on the interviews, while operator centric model is prevalent, the authors expect device centric platforms to narrow down the gaps in developing countries. In particular, Loudon (2016) calls for a focus on SMS as an important platform for M4D services, which could be delivered through various operator and manufacturer-based mobile devices with low cost. Loudon (2016) describes five major design factors to be considered for developing M4D applications and services via SMS: (a) entities that mediate and manage message transmission, (b) SMS billing options, (c) language difference and encoding issues, (d) security, and (e) structured mobile data collection processes.

Examples of M4D projects in remote parts of Africa and India showcase ways in which the low cost and wide reach of mobile technology can be leveraged to design services that help the needy. For example, Mudliar and Donner (2015) focused on a specific mobile technology, interactive voice response (IVR), within the Swara project in India. The IVR allows users to record and listen to messages through cell phones. More than 130,000 calls were received after its launch in 2010, and majority of the recorded content consisted of news reports and complaints over topics such as delays in wage payment and lack of civic amenities. In this way, mobile technology was used as a participatory medium to benefit those who were unable to participate via text or could not afford PCs. Jensen, Iipito, Onwordi, and Mukumbira (2012) investigated how connecting GIS to mobile application to report and map crime could help police officers make better decisions and enhance their response abilities. Through a test in Namibia and interviews of police officers, researchers believed that the mPolicing solution was much more effective and efficient than the traditional system if a full-scale system were to be implemented.

An important aspect of M4D research is a focus on understanding the various uses of mobile media in underdeveloped regions, as they shed light on the affordances of importance and provide insights for designing mobile interventions. Studies often focus on the liberating potential of mobile media for the welfare of women. For example, Maunder, Marsden, and Harper (2011) investigated how a small Xhosa (a South African ethnic group) local community, with women who were working in an institution producing fashion products to develop their occupational skills, use mobile devices to share and obtain information. The researchers installed a system (in an NGO building) that allows people in the community to download and share content through mobile devices within the community. While not all had sufficient time or interest to use the system, among the people who used, the system helped women to share and obtain community information, enhance personal/familial relationship (e.g., interact with and educate children), and enjoy entertainment (e.g., share music or pictures with sound). As another example, Wyche (2017) qualitatively examined Kenyan women's daily mobile usage patterns, and suggested that the concepts of affordability, mobility, and usability, often discussed in ICT4D studies, should be reexamined. For example, while women perceived feature phones as affordable and were generally frugal in terms of their mobile service usage, many of them were interested in new devices and gathered financial resources to purchase the latest smartphones, which speaks to an aspirational outlook. Also, the deep connection and appreciation Kenyan women have for their mobile phones, and the use of mobile phones to primarily communicate with close connections, emerged as other factors that should be considered to design M4D intervention programs. In an M4D study based on a feminist approach, Kwami (2016) examined how Ghanaian women use mobile phones to organize and coordinate transnational trade. Her ethnographic work connected Ghanaian women's attempt to reach personal growth from their marginal status with routinized use of mobile devices. In another ethnographic study, Kibere (2016) demonstrated how youth living in a specific slum area in Kenya (i.e., Kibera) could utilize mobile phones to facilitate communication with others, but primarily to strengthen existing networks in the same low-level income class.

Aside from examining usage among specific groups, M4D research has explored ways in which mobile technology can be applied to improving civic services, mobilization at local levels and business development. For example, in Africa, governments and pressure groups in many countries are realizing that the use of mobile technology can boost citizen participation in governance. A survey of accessibility, skills and attitudes by Ochara and Mawela (2015) revealed that it is practical to enhance civic participation through mobile technology. Even "the socially excluded clusters of individuals were optimistic about the possibilities of mobile technology" (p. 223), especially in organizing at local levels. Mobile technology worked most efficiently by utilizing local connections among citizens. In their study on the effects of mobile technology on economics, Perekwa, Prinsloo, and van Deventer (2016) showed that, for people involved in micro and small enterprises (MSEs) in Zimbabwe, mobile technology benefited them by enhancing their relations with customers, and promoting their productivity and revenues. The researchers also found evidence that mobile technology is creating jobs in various sectors such as mobile banking, mobile advertising, and mobile application development.

Even as M4D projects proliferate around the world, researchers are quick to point out a number of difficulties that pose challenges for designers to tackle. For example, one study identified mobile "utility gaps" between device ownership and actual use of various mobile features in a rural Moroccan community primarily dependent on oral communication. Not only the low levels of literacy, but also various technological (e.g., unstandardized system, poor maintenance), sociolinguistic (e.g., complex language structure, unsupported scripts), and social/cultural barriers (e.g., shared phone use) were found to hinder the usage of mobile phones in such indigenous communities (Dodson, Sterling, & Bennett, 2013).

Carefully calibrating the affordances of mobile technology for the community of interest is an important aspect of M4D work. For example, Donner, Verclas, and Toyama (2008) examined the diversity of projects and approaches in M4D, and suggested future studies to (a) decompose modalities within mobile applications and compare relevant M4D projects for effective evaluation of projects, (b) investigate how micro-level choice (e.g., feature phones vs. smartphones) could have macro-level societal impacts, and (c) better associate content with the context/environment in which the M4D service is being used. Going forth, the HCI community has a lot to offer to M4D projects around the world, in terms of tailoring the design of affordances for particular communities of users.

Concluding Remarks

This chapter has provided an overview of the wide landscape of extant HCI research pertaining to mobile media. The sample of studies and design ideas covered in this chapter simply scratches the surface of the various areas of research, with the goal of providing a beginning researcher a glimpse of this landscape.

Human-computer interaction research with mobile media pertains not only to formal features and affordances of mobile devices but also to the use of mobile applications in a variety of domains. Unlike previous media and communication technologies, mobile media have made haptic modality an integral part of our interaction, with significant innovations in touch interfaces that has introduced a number of affordances, such as swiping to browse images, text entry for messaging, and vibrotactile feedback during game play. Ongoing research emphasizes newer modalities such as speech, gaze, and gesture. Together with touch interfaces, this research holds the promise of innovations in the design and development of multimodal interfaces for mobile media.

Aside from methods of input, HCI research is also concerned with output. With recent improvements in bandwidth and fidelity, mobile media feature advanced graphics and even immersive reality capabilities. Screen sizes have become larger and resolutions sharper. Research has shown that these features have psychological correlates that can shape the manner in which content via mobile media is processed by users. Design efforts will likely yield optimal levels of output characteristics for various combinations of contents and functions, e.g., large screen for improving readability, but smaller screens for systematic processing of content.

A fundamental attribute of mobile media is its mobility, which has inspired a vast amount of research and design pertaining to geographic location of the user. Location-based apps have vastly enhanced the quality of life for the user, with a number of just-in-time aids, such as navigation of physical spaces and coordination with others. Yet, there is more potential to advance design of mobile media, with features that can enhance personal safety of users by providing warnings, improving access to services in the vicinity and so on. Together with the affordances related to recording and streaming live video, location-sharing technology can boost the ability of mobile phones to enhance users' agency by helping them capture ongoing events, including the documentation of police brutalities as witnesses and citizen journalists (Richardson, 2016).

Another important aspect of mobile media is their ubiquity, both in terms of their possession by almost everyone but also in terms of their ability to channel a number of pervasive computing technologies. The potential of mobile media as the proximate interface for ubiquitous computing has inspired design in a number of domains (Sundar, Dou, & Lee, 2013). Applications range from personal organization to accessing e-government services. The domains covered in the chapter speak to the enormous potential of mobile media for not only fulfilling the role of traditional mass media, in terms of providing information and entertainment, but also combining that in interesting ways with more recent social media. In addition, mobile HCI research has shown the way for deploying mobile media for a variety of commercial and civic services, with interfaces that can enhance personal as well as public health and those that can contribute to welfare and development.

In sum, by combining the affordances available in traditional computers and Web-based media (e.g., interactivity, personalization) with newer affordances related to mobility (such as ubiquity and propinquity), mobile media hold the promise of continuing innovations in HCI design and research.

Acknowledgments

The authors wish to thank Ruobing Li and Ruoxu Wang for their research assistance in preparation of this chapter.

References

Au, O. K. C., Su, X., & Lau, R. W. (2014). Linear Dragger: A linear selector for one-finger target acquisition. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)* (pp. 2607–2616). doi:10.1145/2556288.2557096

Banjo, O., Hu, Y., & Sundar, S. S. (2008). Cell phone usage and social interaction with proximate others: Ringing in a theoretical model. *The Open Communication Journal*, *2*, 127–135.

Bayer, J. B., & Campbell, S. W. (2012). Texting while driving on automatic: Considering the frequency-independent side of habit. *Computers in Human Behavior*, *28*, 2083–2090.

Bellman, S., Potter, R. F., Treleaven-Hassard, S., Robinson, J. A., & Varan, D. (2011). The effectiveness of branded mobile phone apps. *Journal of Interactive Marketing*, *25*, 191–200.

Bergé, L.-P., Serrano, M., Perelman, G., & Dubois, E. (2014). Exploring smartphone-based interaction with overview + detail interfaces on 3D public displays. *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '14)* (pp. 125–134). doi:10.1145/2628363.2628374

Birnholtz, J., Fitzpatrick, C., Handel, M., & Brubaker, J. R. (2014). Identity, identification and identifiability: The language of self-presentation on a location-based mobile dating app. *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '14)* (pp. 3–12). doi:10.1145/2628363.2628406

Caballero, M., Chang, T.-R., Menéndez, M., & Occhialini, V. (2010). Behand: Augmented virtuality gestural interaction for mobile phones. *Proceedings of the 12th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '10)* (pp. 451–454). doi:10.1145/1851600.1851704

Cairns, P., Li, J., Wang, W., & Nordin, A. I. (2014). The influence of controllers on immersion in mobile games. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (HCI '14)* (pp. 371–380). doi:10.1145/2556288.2557345

Campbell, S. W., & Kwak, N. (2010). Mobile communication and civic life: Linking patterns of use to civic and political engagement. *Journal of Communication*, *60*, 536–555.

Campbell, S. W., & Kwak, N. (2011a). Mobile communication and civil society: Linking patterns and places of use to engagement with others in public. *Human Communication Research*, *37*, 207–222.

Campbell, S. W., & Kwak, N. (2011b). Political involvement in "mobilized" society: The interactive relationships among mobile communication, network characteristics, and political participation. *Journal of Communication*, *61*, 1005–1024.

Chan-Olmsted, S., Rim, H., & Zerba, A. (2013). Mobile news adoption among young adults examining the roles of perceptions, news consumption, and media usage. *Journalism and Mass Communication Quarterly*, *90*, 126–147.

Cheever, N. A., Rosen, L. D., Carrier, L. M., & Chavez, A. (2014). Out of sight is not out of mind: The impact of restricting wireless mobile device use on anxiety levels among low, moderate and high users. *Computers in Human Behavior*, *37*, 290–297.

Chen, Y., & Pu, P. (2014). HealthyTogether: exploring social incentives for mobile fitness applications. *Proceedings of the Second International Symposium of Chinese CHI* (pp. 25–34). doi:10.1145/2592235.2592240

Chittaro, L., & Sioni, R. (2014). Evaluating mobile apps for breathing training: The effectiveness of visualization. *Computers in Human Behavior*, *40*, 56–63.

Chiu, S. I. (2014). The relationship between life stress and smartphone addiction on Taiwanese university student: A mediation model of learning self-efficacy and social self-efficacy. *Computers in Human Behavior*, *34*, 49–57.

Choe, P., & Schumacher, D. (2015). Influence of different types of vibrations on technical acceptance of a mobile game aiming for hedonic satisfaction. *International Journal of Human-Computer Interaction*, *31*, 33–43.

Chong, A. Y. L., Chan, F. T., & Ooi, K. B. (2012). Predicting consumer decisions to adopt mobile commerce: Cross country empirical examination between China and Malaysia. *Decision Support Systems*, *53*, 34–43.

Church, K., & de Oliveira, R. (2013). What's up with whatsapp? Comparing mobile instant messaging behaviors with traditional SMS. *Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '13)* (pp. 352–361). doi:10.1145/2493190.2493225

Cramer, H., de Juan, P., & Tetreault, J. (2016). Sender-intended functions of emojis in U.S. messaging. *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '16)* (pp. 504–509). doi:10.1145/2935334.2935370

Cuaresma, J., & MacKenzie, I. S. (2014). A comparison between tilt-input and facial tracking as input methods for mobile games. *Games Media Entertainment (GEM, '14 IEEE)* (pp. 1–7). doi:10.1109/GEM.2014.7048080

Cui, Y., Kangas, J., Holm, J., & Grassel, G. (2013). Front-camera video recordings as emotion responses to mobile photos shared within close-knit groups. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)* (pp. 981–990). doi:10.1145/2470654.2466125

Dell, N., Francis, I., Sheppard, H., Simbi, R., & Borriello, G. (2014). Field evaluation of a camera-based mobile health system in low-resource settings. *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '14)* (pp. 33–42). doi:10.1145/2628363.2628366

De Luca, A., Hang, A., von Zezschwitz, E., & Hussmann, H. (2015). I feel like I'm taking selfies all day! Towards understanding biometric authentication on smartphones. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)* (pp. 1411–1414). doi:10.1145/2702123.2702141

de Souza e Silva, A. (2013). Location-aware mobile technologies: Historical, social and spatial approaches. *Mobile Media and Communication*, *1*, 116–121.

Dir, A. L., Cyders, M. A., & Coskunpinar, A. (2013). From the bar to the bed via mobile phone: A first test of the role of problematic alcohol use, sexting, and impulsivity-related traits in sexual hookups. *Computers in Human Behavior*, *29*, 1664–1670.

Dodson, L., Sterling, S., & Bennett, J. (2013). Minding the gaps: Cultural, technical and gender-based barriers to mobile use in oral-language Berber communities in Morocco. *Proceedings of the Sixth International Conference on Information and Communication Technologies and Development (ICTD '13)* (pp. 79–88). doi:10.1145/2516604.2516626

Donner, J., Verclas, K., & Toyama, K. (2008). Reflections on MobileActive08 and the M4D landscape. *Proceedings of the First International Conference on M4D Mobile Communication Technology for Development (M4D '08)* (pp. 73–83). Karlstad, Sweden: Karlstad University.

Dou, X., & Sundar, S. S. (2016). Power of the swipe: Why mobile websites should add horizontal swiping to tapping, clicking, and scrolling interaction techniques. *International Journal of Human-Computer Interaction*, *32*, 352–362.

El-Amrawy, F., & Nounou, M. I. (2015). Are currently available wearable devices for activity tracking and heart rate monitoring accurate, precise, and medically beneficial?. *Healthcare Informatics Research*, *21*, 315–320.

Engl, S., & Nacke, L. E. (2013). Contextual influences on mobile player experience—a game user experience model. *Entertainment Computing*, *4*, 83–91.

Esposito, N., & Lenay, C. (2011). FeelTact: rich tactile feedback for mobile gaming. *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology (ACE '11)* (vol. 71, pp. 1–2). doi:10.1145/2071423.2071511

Fitzpatrick, C., Birnholtz, J., & Gergle, D. (2016). People, places, and perceptions: Effects of location check-in awareness on impressions of strangers. *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices & Services (MobileHCI '16)* (pp. 295–305). doi:10.1145/2935334.2935369

Forgays, D. K., Hyman, I., & Schreiber, J. (2014). Texting everywhere for everything: Gender and age differences in cell phone etiquette and use. *Computers in Human Behavior*, *31*, 314–321.

Frith, J. (2013). Turning life into a game: Foursquare, gamification, and personal mobility. *Mobile Media and Communication*, *1*, 248–262.

Furió, D., GonzáLez-Gancedo, S., Juan, M.-C., Seguí, I., & Rando, N. (2013). Evaluation of learning outcomes using an educational iPhone game vs. traditional game. *Computers and Education*, *64*, 1–23.

Gao, L., Waechter, K. A., & Bai, X. (2015). Understanding consumers' continuance intention towards mobile purchase: A theoretical framework and empirical study–A case of China. *Computers in Human Behavior*, *53*, 249–262.

Girouard, A., Lo, J., Riyadh, M., Daliri, F., Eady, A. K., & Pasquero, J. (2015). One-handed bend interactions with deformable smartphones. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)* (pp. 1509–1518). doi:10.1145/2702123.2702513

Goh, K. Y., Chu, J., & Wu, J. (2015). Mobile advertising: An empirical study of temporal and spatial differences in search behavior and advertising response. *Journal of Interactive Marketing*, *30*, 34–45.

Göllner, S., Hurtienne, J., Gollner, U., & Naumann, A. B. (2011). Mobile technology keeping people with dementia independent and socially active. *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '11)* (pp. 593–598). doi:10.1145/2037373.2037466

Gonzales, A. L. (2014). Text-based communication influences self-esteem more than face-to-face or cellphone communication. *Computers in Human Behavior*, *39*, 197–203.

Greenberg, S., & Neustaedter, C. (2013). Shared living, experiences, and intimacy over video chat in long distance relationships. In C. Neustaedter, S. Harrison, & A. Sellen (Eds.), *Connecting families: The impact of new communication technologies on domestic life* (pp. 37–53). London: Springer.

Hadlington, L. J. (2015). Cognitive failures in daily life: Exploring the link with Internet addiction and problematic mobile phone use. *Computers in Human Behavior*, *51*, 75–81.

Häkkilä, J. R., Posti, M., Schneegass, S., Alt, F., Gultekin, K., & Schmidt, A. (2014). Let me catch this!: Experiencing interactive 3D cinema through collecting content with a mobile phone. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)* (pp. 1011–1020). doi:10.1145/2556288.2557187

Heinrichs, F., Schreiber, D., & Schöning, J. (2011). The hybrid shopping list: Bridging the gap between physical and digital shopping lists. *Proceedings of the 13th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '11)* (pp. 251–254). doi:10.1145/2037373.2037411

Henze, N., Rukzio, E., & Boll, S. (2012). Observational and experimental investigation of typing behaviour using virtual keyboards for mobile devices. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)* (pp. 2659–2668). doi:10.1145/2207676.2208658

Hjorth, L., & Pink, S. (2014). New visualities and the digital wayfarer: Reconceptualizing camera phone photography and locative media. *Mobile Media and Communication*, *2*, 40–57.

Hofmann, S., & Mosemghvdlishvili, L. (2014). Perceiving spaces through digital augmentation: An exploratory study of navigational augmented reality apps. *Mobile Media and Communication*, *2*, 265–280.

Humphreys, L. (2013). Mobile social media: Future challenges and opportunities. *Mobile Media and Communication*, *1*, 20–25.

Humphreys, L., & Liao, T. (2011). Mobile geotagging: Reexamining our interactions with urban space. *Journal of Computer-Mediated Communication*, *16*, 407–423.

Hwangbo, H., Yoon, S. H., Jin, B. S., Han, Y. S., & Ji, Y. G. (2013). A study of pointing performance of elderly users on smartphones. *International Journal of Human-Computer Interaction*, *29*, 604–618.

Jarusriboonchai, P., Malapaschas, A., & Olsson, T. (2016). Design and evaluation of a multiplayer mobile game for icebreaking activity. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)* (pp. 4366–4377). doi: 10.1145/2858036.2858298

Jensen, K. L., Iipito, H. N., Onwordi, M. U., & Mukumbira, S. (2012). Toward an mPolicing solution for Namibia: Leveraging emerging mobile platforms and crime mapping. In *Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference (SATCSIT '12)* (pp. 196–205). doi:10.1145/2389836.2389860

Kang, J.-Y. M., Mun, J. M., & Johnson, K. K. P. (2015). In-store mobile usage: Downloading and usage intention toward mobile location-based retail apps. *Computers in Human Behavior*, *46*, 210–217.

Kang, S.-H., & Watt, J. H. (2013). The impact of avatar realism and anonymity on effective communication via mobile devices. *Computers in Human Behavior*, *29*, 1169–1181.

Karippacheril, T. G., Nikayin, F., de Reuver, M., & Bouwman, H. (2013). Serving the poor: Multisided mobile service platforms, openness, competition, collaboration and the struggle for leadership. *Telecommunications Policy*, *37*, 24–34.

Kibere, F. N. (2016). The paradox of mobility in the Kenyan ICT ecosystem: An ethnographic case of how the youth in Kibera slum use and appropriate the mobile phone and the mobile internet. *Information Technology for Development*, *22*, 47–67.

Kim, C., Mirusmonov, M., & Lee, I. (2010). An empirical examination of factors influencing the intention to use mobile payment. *Computers in Human Behavior*, *26*, 310–322.

Kim, H., & Lingel, J. (2016). Working through paradoxes: Transnational migrants' urban learning tactics using locative technology. *Mobile Media and Communication*, *4*, 221–236.

Kim, K. J., Park, E., & Sundar, S. S. (2012). IPS vs. AMOLED: Effects of panel type on smartphone users' viewing and reading experience. *Human Centric Technology and Service in Smart Space*, *182*, 77–84.

Kim, K. J., & Sundar, S. S. (2014). Does screen size matter for smartphones? Utilitarian and hedonic effects of screen size on smartphone adoption. *Cyberpsychology, Behavior, and Social Networking*, *17*, 466–473.

Kim, K. J., & Sundar, S. S. (2016). Mobile persuasion: Can screen size and presentation mode make a difference to trust?. *Human Communication Research*, *42*, 45–70.

Kim, K. J., Sundar, S. S., & Park, E. (2011). The effects of screen-size and communication modality on psychology of mobile device users. In *CHI '11 Extended Abstracts on Human Factors in Computing Systems* (pp. 1207–1212). doi:10.1145/1979742.1979749

Kim, S. J., Wang, R. J. H., & Malthouse, E. C. (2015). The effects of adopting and using a brand's mobile application on customers' subsequent purchase behavior. *Journal of Interactive Marketing*, *31*, 28–41.

Kwami, J. D. (2016). Development from the margins? Mobile technologies, transnational mobilities, and livelihood practices among Ghanaian women traders. *Communication, Culture and Critique*, *9*, 148–168.

Lam, C. (2013). The efficacy of text messaging to improve social connectedness and team attitude in student technical communication projects: An experimental study. *Journal of Business and Technical Communication*, *27*, 180–208.

Lee, S., Bang, H., & Sundar, S. S. (2014). Social interaction in mobile games: Priming opponents' presence and haptic feedback. *Proceedings of HCI Korea (HCIK '15)* (pp. 104–108). Seoul, South Korea: Hanbit Media.

Lee, S., Kim, K. J., & Sundar, S. S. (2015). Customization in location-based advertising: Effects of tailoring source, locational congruity, and product involvement on ad attitudes. *Computers in Human Behavior*, *51*, 336–343.

Lepp, A., Barkley, J. E., & Karpinski, A. C. (2014). The relationship between cell phone use, academic performance, anxiety, and satisfaction with life in college students. *Computers in Human Behavior*, *31*, 343–350.

Li, J., Lepp, A., & Barkley, J. E. (2015). Locus of control and cell phone use: Implications for sleep quality, academic performance, and subjective well-being. *Computers in Human Behavior*, *52*, 450–457.

Li, K., & Du, T. C. (2012). Building a targeted mobile advertising system for location-based services. *Decision Support Systems*, *54*, 1–8.

Li, M., Scharf, K. M., & Kobbelt, L. (2016). MobileVideoTiles: Video display on multiple mobile devices. *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct (MobileHCI '16)* (pp. 621–626). doi:10.1145/2957265.2961826

Li, R., Zhang, B., Sundar, S. S., & Duh, H. B. L. (2013). Interacting with augmented reality: How does location-based AR enhance learning?. In P. Kotzé, G. Marsden, G. Lindgaard, J. Wesson, & M. Winckler (Eds.), *Human-Computer Interaction—INTERACT 2013* (pp. 616–623). Berlin: Springer.

Liébana-Cabanillas, F., Sánchez-Fernández, J., & Muñoz-Leiva, F. (2014). Antecedents of the adoption of the new mobile payment systems: The moderating effect of age. *Computers in Human Behavior*, *35*, 464–478.

Lin, T. T., Younbo, J., & Sim, C. (2015). Towards an understanding of intention to use mobile videos: Impression management, perceived facilitation, and social norms. *Mobile Media and Communication*, *3*, 106–124.

Ling, R., & Lai, C.-H. (2016). Microcoordination 2.0: Social coordination in the age of smartphones and messaging apps. *Journal of Communication*, *66*, 834–856.

Loudon, M. (2016). A platform studies approach to the role of technology in the ICTD ecosystem: The SMS in M4D interventions. *Information Technology for Development*, *22*, 7–25.

Lu, C., Chang, M., Huang, E., & Chen, C.-W. (2015). The effect of story in mobile educational game. In K. R. Huang (Ed.), *Ubiquitous learning environments and technologies* (pp. 55–76). Berlin, Germany: Springer.

Maunder, A., Marsden, G., & Harper, R. (2011). Making the link—providing mobile media for novice communities in the developing world. *International Journal of Human-Computer Studies*, *69*, 647–657.

Mudliar, P., & Donner, J. (2015). Experiencing interactive voice response (IVR) as a participatory medium: The case of CGNet Swara in India. *Mobile Media and Communication*, *3*, 366–382.

Niu, J., Liu, Y., Lin, J., Zhu, L., & Wang, K. (2014). Stroke++: A new Chinese input method for touch screen mobile phones. *International Journal of Human-Computer Studies*, *72*, 440–450.

Ochara, N. M., & Mawela, T. (2015). Enabling social sustainability of e-participation through mobile technology. *Information Technology for Development*, *21*, 205–228.

O'Hara, K., Mitchell, A. S., & Vorbau, A. (2007). Consuming video on mobile devices. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)* (pp. 857–866). doi:10.1145/1240624.1240754

Okazaki, S., & Mendez, F. (2013). Exploring convenience in mobile commerce: Moderating effects of gender. *Computers in Human Behavior*, *29*, 1234–1242.

Oppermann, L., Blum, L., & Shekow, M. (2016). Playing on AREEF: Evaluation of an underwater augmented reality game for kids. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '16)* (pp. 330–340). doi:10.1145/2935334.2935368

Park, E. K., & Sundar, S. S. (2015). Can synchronicity and visual modality enhance social presence in mobile messaging? *Computers in Human Behavior*, *45*, 121–128.

Park, T., & Salvendy, G. (2012). Emotional factors in advertising via mobile phones. *International Journal of Human-Computer Interaction*, *28*, 597–612.

Park, W., & Han, S. H. (2014). An analytical approach to creating multitouch gesture vocabularies in mobile devices: A case study for mobile web browsing gestures. *International Journal of Human-Computer Interaction*, *30*, 126–141.

Perekwa, G. B., Prinsloo, T., & van Deventer, J. (2016). The impact of mobile technology on micro and small enterprises in Zimbabwe in the post-hyperinflation economic era. *The African Journal of Information Systems*, *8*(3). 45–66.

Pescher, C., Reichhart, P., & Spann, M. (2014). Consumer decision-making processes in mobile viral marketing campaigns. *Journal of Interactive Marketing*, *28*, 43–54.

Poppinga, B., Shirazi, A. S., Henze, N., Heuten, W., & Boll, S. (2014). Understanding shortcut gestures on mobile touch devices. *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services (MobileHC I '14)* (pp. 173–182). doi:10.1145/2628363.2628378

Procyk, J., Neustaedter, C., Pang, C., Tang, A., & Judge, T. K. (2014). Exploring video streaming in public settings: Shared geocaching over distance using mobile video chat. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)* (pp. 2163–2172). doi:10.1145/2556288.2557198

Quinn, S., & Oldmeadow, J. (2013). The martini effect and social networking sites: Early adolescents, mobile social networking and connectedness to friends. *Mobile Media and Communication*, *1*, 237–247.

Raptis, D., Tselios, N., Kjeldskov, J., & Skov, M. B. (2013). Does size matter? Investigating the impact of mobile phone screen size on users' perceived usability, effectiveness and efficiency. *Proceedings of the 15th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '13)* (pp. 127–136). doi:10.1145/2493190.2493204

Richardson, A. V. (2016). Bearing witness while black: Theorizing African American mobile journalism after Ferguson. *Digital Journalism*, *5*(6), 1–26.

Roberts, J. A., & David, M. E. (2016). My life has become a major distraction from my cell phone: Partner phubbing and relationship satisfaction among romantic partners. *Computers in Human Behavior*, *54*, 134–141.

Romano, M., Paolino, L., Tortora, G., & Vitiello, G. (2014). The tap and slide keyboard: A new interaction method for mobile device text entry. *International Journal of Human-Computer Interaction*, *30*, 935–945.

Rooksby, J., Rost, M., Morrison, A., & Chalmers, M. (2015). Pass the ball: Enforced turn-taking in activity tracking. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)* (pp. 2417–2426). doi:10.1145/2702123.2702577

Rost, M., Kitsos, C., Morgan, A., Podlubny, M., Romeo, P., Russo, E., & Chalmers, M. (2016). Forget-me-not: History-less mobile messaging. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 1904–1908). doi:10.1145/2858036.2858347

Rozado, D., Moreno, T., San Agustin, J., Rodriguez, F. B., & Varona, P. (2015). Controlling a smartphone using gaze gestures as the input mechanism. *Human–Computer Interaction*, *30*, 34–63.

Ruiz, J., Li, Y., & Lank, E. (2011). User-defined motion gestures for mobile interaction. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)* (pp. 197–206). doi:10.1145/1978942.1978971

Salehan, M., & Negahban, A. (2013). Social networking on smartphones: When mobile phones become addictive. *Computers in Human Behavior*, *29*, 2632–2639.

Schepman, A., Rodway, P., Beattie, C., & Lambert, J. (2012). An observational study of undergraduate students' adoption of note-taking software. *Computers in Human Behavior*, *28*, 308.

Shin, D.-H., Shin, Y.-J., Choo, H., & Beom, K. (2011). Smartphones as smart pedagogical tools: Implications for smartphones as u-learning devices. *Computers in Human Behavior*, *27*, 2207–2214.

Shirali-Shahreza, S., Penn, G., Balakrishnan, R., & Ganjali, Y. (2013). Seesay and hearsay captcha for mobile interaction. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)* (pp. 2147–2156). doi:10.1145/2470654.2481295

Shuib, L., Shamshirband, S., & Ismail, M. H. (2015). A review of mobile pervasive learning: Applications and issues. *Computers in Human Behavior*, *46*, 239–244.

Smith, A. (2015). *U.S. smartphone use in 2015*. Washington, DC: Pew Research Center. Retrieved from http://www.pewinternet.org/2015/04/01/us-smartphone-use-in-2015/

Sterling, G. (2016, April 4). Nearly 80 percent of social media time now spent on mobile devices. *Marketing Land*. Retrieved from http://marketingland.com/facebook-usage-accounts-1–5-minutes-spent-mobile-171561

Strassberg, D. S., Rullo, J. E., & Mackaronis, J. E. (2014). The sending and receiving of sexually explicit cell phone photos ("Sexting") while in high school: One college's students' retrospective reports. *Computers in Human Behavior*, *41*, 177–183.

Su, Y.-S., Chiang, W.-L., Lee, C.-T. J., & Chang, H.-C. (2016). The effect of flow experience on player loyalty in mobile game application. *Computers in Human Behavior*, *63*, 240–248.

Sundar, S. S., Bellur, S., Oh, J., Xu, Q., & Jia, H. (2014). User experience of on-screen interaction techniques: An experimental investigation of clicking, sliding, zooming, hovering, dragging, and flipping. *Human–Computer Interaction*, *29*, 109–152. doi:10.1007/978-3-642-40480-1_51

Sundar, S. S., Dou, X., & Lee, S. (2013). Communicating in a Ubicomp world: Interaction rules for guiding design of mobile interfaces. In P. Kotzé, G. Marsden, G. Lindgaard, J. Wesson & M. Winkler (Eds.), *Lecture notes in computer science: Vol. 8118. Human-computer interaction—INTERACT 2013* (pp. 730–747). Berlin: Springer.

Tang, H., Liao, S. S., & Sun, S. X. (2013). A prediction framework based on contextual data to support mobile personalized marketing. *Decision Support Systems*, *56*, 234–246.

Tong, T., Guana, V., Jovanovic, A., Tran, F., Mozafari, G., Chignell, M., & Stroulia, E. (2015). Rapid deployment and evaluation of mobile serious games: A cognitive assessment case study. *Procedia Computer Science*, *69*, 96–103.

Tossell, C. C., Kortum, P., Shepard, C., Barg-Walkow, L. H., Rahmati, A., & Zhong, L. (2012). A longitudinal study of emoticon use in text messaging from smartphones. *Computers in Human Behavior*, *28*, 659–663.

Tu, H., & Ren, X. (2013). Optimal entry size of handwritten Chinese characters in touch-based mobile phones. *International Journal of Human-Computer Interaction*, *29*, 1–12.

Tu, H., Ren, X., Tian, F., & Wang, F. (2014). Evaluation of flick and ring scrolling on touch-based smartphones. *International Journal of Human-Computer Interaction*, *30*, 643–653.

Van Damme, K., Courtois, C., Verbrugge, K., & De Marez, L. (2015). What's APPening to news? A mixed-method audience-centred study on mobile news consumption. *Mobile Media and Communication*, *3*, 196–213.

Varshney, U. (2014). Mobile health: Four emerging themes of research. *Decision Support Systems*, *66*, 20–35.

Veras, R., Paluka, E., Chang, M. W., Tsang, V., Shein, F., & Collins, C. (2014). Interaction for reading comprehension on mobile devices. *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '14)* (pp. 157–161). doi:10.1145/2628363.2628387

Wacker, P., Kreutz, K., Heller, F., & Borchers, J. (2016). Maps and Location: Acceptance of modern interaction techniques for audio guides. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)* (pp. 1067–1071). doi:10.1145/2858036.2858189

Weilenmann, A., Säljö, R., & Engström, A. (2014). Mobile video literacy: Negotiating the use of a new visual technology. *Personal and Ubiquitous Computing*, *18*, 737–752.

Weiss, A. S. (2013). Exploring news apps and location-based services on the smartphone. *Journalism and Mass Communication Quarterly*, *90*, 435–456.

Wijetunga, D. (2014). The digital divide objectified in the design: Use of the mobile telephone by underprivileged youth in Sri Lanka. *Journal of Computer-Mediated Communication*, *19*, 712–726.

Wolf, K., & Henze, N. (2014). Comparing pointing techniques for grasping hands on tablets. *Proceedings of the 16th International Conference on Human-Computer Interaction with Mobile Devices & Services (MobileHCI '14)* (pp. 53–62). doi:10.1145/2628363.2628371

Wyche, S. (2017). Exploring women's everyday mobile phone experiences in Nairobi, Kenya. *Interacting with Computers*, *29*, 391–402. doi:10.1093/iwc/iww028

Yang, R., Shin, E., Newman, M. W., & Ackerman, M. S. (2015). When fitness trackers don't "fit": End-user difficulties in the assessment of personal tracking device accuracy. *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp'15)* (pp. 623–634). doi:10.1145/2750858.2804269

Yang, S., Lu, Y., Gupta, S., Cao, Y., & Zhang, R. (2012). Mobile payment services adoption across time: An empirical study of the effects of behavioral beliefs, social influences, and personal traits. *Computers in Human Behavior*, *28*, 129–142.

Yumak, Z., & Pu, P. (2013). Survey of sensor-based personal wellness management systems. *BioNanoScience*, *3*, 254–269.

Zhang, L., Zhu, J., & Liu, Q. (2012). A meta-analysis of mobile commerce adoption and the moderating effect of culture. *Computers in Human Behavior*, *28*, 1902–1911.

Zhou, J., Rau, P. L. P., & Salvendy, G. (2014). Older adults' text entry on smartphones and tablets: Investigating effects of display size and input method on acceptance and performance. *International Journal of Human-Computer Interaction*, *30*, 727–739.

28

Natural Human-Robot Interaction

Yasser Mohammad

Introduction

Naturalness in this chapter refers to a quality of the interaction between the robot and the human that is becoming increasingly important with the advent of service and social robots. Naturalness is a multifaceted concept and this chapter tries to address its multitude of meanings and incarnations in human-robot interaction (HRI) as evident in current research.

This chapter has been designed to give a wide view of the research in this area. The following section introduces the concept of natural HRI with its different methods and interrelated concepts, and tries to differentiate between it and related concepts like humanlike interaction and human-centered robotics. We then describe different modalities of natural HRI and provide an overview of some of the research in these modalities, and describe two alternative techniques to achieve natural HRI (the engineering and the machine-learning approaches). Finally, the chapter presents two examples of interaction corpora that can be used for research in the area of natural HRI.

What is Natural HRI?

Valli (2008) defines natural interaction "in terms of experience: people naturally communicate through gestures, expressions, movements, and discover the world by looking around and manipulating physical stuff; the key assumption here is that they should be allowed to interact with technology as they are used to interact with the real world in everyday life, as evolution and education taught them to do." This view of naturalness focuses on giving the robot the ability to respond to people's normal modalities of interaction including gestures, motions and—although it is not mentioned—verbal dialogue. Notice that this definition focuses on the robot's ability to *understand* natural means of interaction rather than being necessarily able to produce similar kinds of behavior. We call this kind of robot a *naturally controllable robot*. This capacity to understand natural communication acts will only become more

The Wiley Handbook of Human Computer Interaction, Volume 2, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

important with time as robots move beyond their current niches in factories to live with technology naïve users at homes, offices, hospitals, and schools. It is also very important for robots working in stressful environment like disaster sites where people are more likely to prefer using their own communicative behaviors to control robots instead of artificial interfaces.

Another definition of natural HRI (at least for humanoid robots) focuses on the ability of the robot to generate *expected* behavior that reveals its internal state accurately to its human interlocutors. Expected behavior is not necessary humanlike behavior but it can be so. Consider a pet robot designed to look like a dog (e.g. Sony's AIBO). Such a robot will be expected to behave more like a pet than like a person. It would be expected to learn over time, like normal animals. For humanoid robots, this kind of naturalness translates in many cases to human likeness. By virtue of appearance, humanoids are expected by most users to behave in a similar way to humans, to invoke our bias toward anthropomorphism. Bisio et al. (2014) have shown that motor contagion happens when robots move in a biologically plausible manner but not when they violate the expectations of biological motion. In this experiment, participants observed a humanoid robot and a human agent move their hands into a prespecified final position or put an object into a container at various velocities. Their movements, both in the object- and non-object-directed conditions, were characterized by either a smooth/curvilinear or a jerky/segmented trajectory. These trajectories were covered with biological or nonbiological kinematics (the latter only by the humanoid robot). After action observation, participants were requested to either reach the indicated final position or to transport a similar object into another container. Results showed that motor contagion appeared for both the interactive partner except when the humanoid robot violated the biological laws of motion. Chartrand and Bargh (1999) experimentally showed that behavioral mimicry (the chameleon effect) has a significant effect on the interaction and increases empathy towards the interaction partner. Human-robot interaction studies documented similar effects. Riek, Paul and Robinson (2010) used real-time head-gesture mimicry to improve rapport between the human and a robot and Lee et al. (2004) programmed a robotic penguin to nod in return to detected nods in order to consolidate the backchannel communication in a natural way. What this study and similar ones reveal is that many of the same responses to human behavior appear in human-humanoid interactions and provide a basis of shared understanding that should be utilized by the robot if the interaction is to go on naturally. Robots that can behave in this expected manner (either humanlike or not) are called naturally behaving robots (in contrast to naturally controllable robots discussed earlier).

The two faces of natural robots mentioned above (natural controllability and natural behavior) can be combined to produce easy-to-control robots that can produce natural behavior. It may not be clear what is the expected natural behavior of the robot of a given form at a given context. A possible solution to this problem is to adopt a definition of naturalness in HRI that relates it directly to its effect on human partners of the robot. Natural behavior, hence, can be defined as the robotic behavior that induces minimal negative emotions (e.g. cognitive load, stress, and frustration) and maximizes positive emotions (e.g. engagement and fun). This definition of natural HRI can be used to devise an objective measure of robot's behavior that is mostly independent from the task or context based on signal analysis of the physiological correlates of psychological state like skin conductance, heart-rate variability, and so forth. It is also possible to rely on subjective evaluation of the robot's behavior to

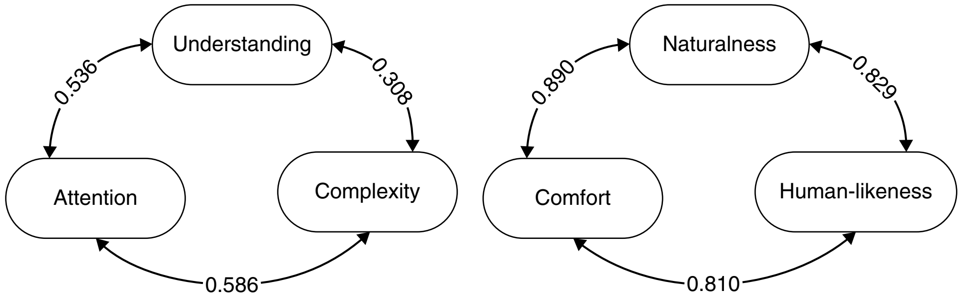


Figure 28.1 Correlation between dimensions of evaluation when interacting with robots using different gaze-control mechanisms.

measure this aspect of naturalness. One challenge to this approach is the difficulty in ascertaining psychological state changes based on physiological signals in normal situations without inducing extreme conditions. We will discuss a specific technique for handling this problem later in this chapter.

Mohammad, Nishida and Okada (2010) compared the subjective evaluations of 48 participants after watching videos of a Robovie II humanoid robot listening to an explanation by a human using two different gaze control mechanisms. Figure 28.1 shows the results of correlation analysis between different dimensions of subjective evaluation in this experiment. It is clear that two distinct clusters appear where naturalness is correlated highly with human likeness (naturally behaving robots) and comfort. This result highlights the interplay between the second and third definitions of naturalness discussed in this section (natural behavior and behavior inducing positive emotions) at least for this task.

Natural HRI Modalities

Human communication utilizes several interaction channels. As robots become more social, they are expected to understand and use the same channels. This section discusses some of the approaches roboticists utilized for giving robots such abilities. The discussions in this section are not designed to be exhaustive but to provide a general overview of the field. Interested readers are advised to use the references at the end of this chapter to gain more information about any of the studies reported here.

Natural nonverbal interaction

Nonverbal communication is responsible for 70% of communication during human-human face-to-face interactions. There are many kinds of nonverbal behavior used during face to face interaction. Explicit and implicit protocols appear with different ratios in each of these channels. Researchers in human-human interaction usually classify these into (Argyle, 2001):

- gaze (and pupil dilation);
- spatial behavior;

- gestures and other bodily movements;
- nonverbal vocalization;
- posture;
- facial expression;
- bodily contact;
- clothes and other aspects of appearance;
- smell.

This list was ordered, roughly, by applicability to HRI and especially suitability for learning the corresponding nonverbal behavior using the machine-learning approach presented in this chapter.

Smell is rarely used in HRI because of the large difference in form factor between most robots and humans, which is expected to lead to different perception of smells than in the human case. Clothes and other aspects of appearance are important nonverbal signaling channels during human-human interaction and differences in appearance between robots is known to affect how people respond to them and the “uncanny valley” is a clear example of such an effect (Bartneck, Kanda, & Norihiro Hagita, 2009). Mori, MacDorman, and Kageki (2012) proposed the uncanny valley hypothesis, which posits that increased robotic anthropomorphism increases the positivity of humans’ emotional response to them up to a point at which the similarity between the robot and a human becomes too close, so that further increases in human likeness lead to decreased likability (to a level even worse than clearly mechanical robotic forms) followed by another increase in likability when the robot becomes indistinguishable from humans and further on. If human likeness is plotted against likability, according to this hypothesis, a valley appears, which is what gives the hypothesis its name. The hypothesis further predicts that robotic motion exaggerates the whole curve leading to even faster increase and decrease of likability with human likeness.

Industrial robots used to work in cages with no contact with human workers, to ensure the safety of humans. Direct bodily contact is becoming increasingly important in HRI in applications such as haptic interfaces, cooperative material handling, power extension using robots, rehabilitation and physical training, and entertainment. This plethora of applications provide new challenges for HRI that were never faced in HCI applications including cognitive issues (e.g. perception, or mental modeling of motion), and physical issues (e.g. safety and dependability).

Facial expressiveness is still far from being humanlike for most robots with the exception of Germinoinds that use artificial skin and are usually modeled after a specific person. This usually entails a change in the situation from interaction to detection of a human’s facial expressions and its associated meanings (e.g. in affective computing).

Posture is an important nonverbal signal. It conveys information about the internal state of the poser and can be used to analyze power distribution in the interaction situation (Argyle, 2001).

Nonverbal vocalization usually comes with verbal behavior. Several kinds of nonverbal vocalization have been studied in human-human interactions, including prosody, intonation, dramatic pauses, and changes in tone and tempo.

Gestures and other body movements can be used both during implicit and explicit protocols. Iconic gestures can be used explicitly to pass information between communication partners. Spontaneous gestures accompany speech in many cases and provide nonverbal connotations to this predominantly verbal channel. A robot's ability to encode gestures depends on the degrees of freedom it has in its hands and other parts of the body. This may limit the robot's ability to produce gestures in many cases, yet gesture recognition and appropriate responses to human gestures are widely used in HRI and will be discussed in more detail later in this chapter.

Spatial behavior appears in human-human interactions in many forms including distance management and body alignment. As with posture, relative spatial orientation and distance provide valuable information about power distribution in the interaction as well as an indication of the personal relation between different partners.

Gaze is to be one of the most useful nonverbal behaviors both in human-human and human-robot interactions. Infants prefer to look at faces that engage them in mutual gaze from birth and healthy babies show enhanced neural processing of direct gaze from an early age (Farroni, Csibra, Simion, & Johnson, 2002). Shared attention is also signaled through this communication channel and is one of the building blocks in the development of a theory of mind (ToM) in children according to modularists (Baron-Cohen, Leslie, & Frith, 1985).

Naturally Controllable Robots

The first definition of naturalness in HRI considered in this chapter was natural controllability, which refers to the robot's ability to understand and respond appropriately to the intentions of human operators conveyed through natural interaction modalities. One such modality that is used intensively in HRI research is gesture. The ability to detect a single gesture like pointing can enhance the ease with which the robot can be controlled. For example, Waldherr, Romero, and Thrun (2000) developed a gesture interface for controlling a mobile robot. A camera is used to detect body and arm orientations. This data is fed to either a template based system or a neural network to detect gestures that trigger different actions from the robot. The system was tested in an interactive cleanup task in which a human operator guides the robot to locations that need cleaning up and instructs it (through predefined gestures) to pick up trash.

Gesture interfaces

Gesture interfaces can be combined with other verbal and nonverbal interfaces to produce multimodal interactions. For example, Stiefelhagen et al. (2004) combined dialogue, head-pose detection and pointing gesture detection to provide a natural interaction interface for ARMAR, a humanoid robot with two arms and 23 degrees of freedom.

Complex dynamic and parameterized gestures can be detected from videos and point cloud data. For example, Bennewitz, M., Axenbeck, T., Behnke, S., and Burgard (2008) used HMMs to detect a variety of complex and parameterized gestures. The gesture is first decomposed into phases and a HMM is trained to recognize each of these phases. Composed HMMs are then created that model the transition between individual phase HMMs. Once a specific phase is recognized, we estimate the parameter of a gesture such as the target of a pointing gesture. Rautaray and Agrawal (2015) provide a survey of vision-based gesture recognition for HCI in general.

With the advent of low-cost RGB-D cameras like Kinect, gesture recognition research for robotics applications shifted toward detecting more complex gestures based on these new sensors. For example, Cicirelli, Attolico, Guaragnella, and D'Orazio (2015) used three Kinect cameras to control a remote robot using predefined gestures. The system is based on skeletal tracking using OpenNI, which limits the types of gestures that can be learned, yet, it was successful in allowing naïve users to teleoperate a robot in a navigation task.

One common feature of these studies is that a predefined set of gestures is used for interacting with the robot. A more natural interface would allow humans to use whatever gestures they may use with other humans in interacting with the robot. This can be achieved by careful analyzing gesture communication in some context and implementing gesture recognition routines of all gestures involved in the robot. This solution, though, is not scalable and will require human involvement whenever a new context or a new culture is concerned. A more scalable approach is to design a learning system that allows the robot to discover gestures and understand their communicative value by watching human-human interactions in the real world. Systems employing this approach are discussed later in this chapter.

Natural language HRI

Verbal interaction is one of the most information-rich modalities for human-human communication and provides a natural medium to transfer commands and receive feedback from robot. Verbal communication usually takes the form of a dialogue, which makes spoken dialogue management a major challenge for HRI based on spoken language.

The simplest verbal communication mode in HRI is direct verbal commands. Examples of such systems include MAIA, which can carry objects from place to place following direct verbal commands, and RHINO, which provides exhibit tours in a museum when verbally ordered to do so (Burgard et al., 1998). These early system employed a rigid interaction scenario in which the robot is passively understanding the human command providing a naturally controllable robot. Jijo-2 was one of the early systems that provided more natural verbal communication by employing a simple spoken dialog system (SDS) (Matsui et al., 1999).

The most widely used—and the simplest—SDSs are state-based systems that model the dialogue using state machines. Utterances are used to switch between states and provide information to fill specific slots attached with each state that can then be used to execute external commands and control the behavior of the robot. Even though these systems provide rudimentary interactivity, the flexibility of the interaction remains minimal. The main advantage of state-based SDSs is the predictability of user utterances, which not only simplifies the design of the dialogue system but aids in improving the accuracy of speech recognition.

A more advanced kind of SDSs uses frames that represent tasks and subtasks. User utterances are used to fill information slots in these tasks without having to follow a predefined state transition graph. This allows for higher flexibility compared with state-based systems leading to less predictable user utterances and increased complexity in dialogue management.

Plan-based SDSs are designed to recognized to the user's plan and the role of the robot in it. Utterances are employed to incrementally adjust the robot's initial perception of the plan. These systems are the most flexible of the four alternatives discussed so far allowing shortcuts based on identification of information throughout the dialogue (Spiliotopoulos, Androutsopoulos, & Spyropoulos, 2001).

Recently, research in human-robot dialogue started to focus not only on the verbal aspects of the dialogue per se as described in the systems mentioned above but on the contextual information surrounding the dialogue. For example, Chai et al. (2014) studied dialogue grounding based on ideas from Clark (1996). Dialogue is modeled as a series of *contributions*, each consists of two phases: a *presentation* phase and an *acceptance* phase. During presentation, the speaker presents some utterance to the addressee (either of which can be the robot) and the addressee later confirms understanding of this utterance in the acceptance phase. This provides a mechanism for second order belief about the piece of information or command that was given in the presentation phase. Accumulation of these believes generate the common ground between the robot and the human. Research in situated dialogue is still in its infancy and much work is still needed to bridge the perceptual gap between the human and the robot, allowing for more implicit common ground formation and maintenance.

Multimodal natural HRI

Gesture and spoken language provide but two natural modalities for interacting with robots. Human communication usually employs multiple modalities at once. We usually gesture while speaking and Argyle (2001) speculates that these speech-accompanied gestures are generated along with speech from the same process encoding the idea to be communicated. It is then natural to expect humans to use multiple modalities to communicate with robots as well.

Kollar et al. (2012) proposed a multimodal HRI system that combines verbal understanding through speech recognition and Bayesian inference with RGB-D-based gesture recognition in a receptionist robot and showed that the robot can successfully interact with people 75% of the time when they are primed with its capabilities and only 57% of the time when they are not familiar with the robot's capabilities. Swaminathan and Sridharan (2011) proposed a robotic system that utilizes both visual and verbal information to incrementally learn multimodal models of objects. Perzanowski, Schultz, Adams, Marsh, and Bugajska (2001) built a multimodal interaction system that combined gesture recognition, verbal communication and PDA-based direct commands. The system was designed to allow the users more freedom in focusing in the tasks at hand instead of the communication with the robot. Gestures were employed in the system to provide either estimates of distance / size by holding the hands apart or indicating a direction by tracing a line with the hand in air. These limited gestures were employed due to the limitations of the robot's vision system.

Gorostiza et al. (2006) developed an architecture called automatic deliberative, which employs an emotional control system on a robot called Maggie, which was design to interact in peer-to-peer situations with humans. The system can combine tactile, visual and voice modalities based on markup languages that represent the information gathered through all of these modalities as well as robot control strategies. Stiefelhagen et al. (2004) developed a system that can integrate pointing gesture recognition, head pose analysis, verbal communication and dialogue management to implement a personal companion robot that interacts with people in a kitchen environment.

Naturally Behaving Robots

Understanding the intentions of humans is the main challenge facing naturally controllable robots like the ones described in the previous section. Naturally behaving robots face this same problem combined with the problem of conveying their intentions to humans using natural means of communication.

Mohammad & Nishida (2007) designed an experiment to evaluate behavior naturalness of a miniature nonhumanoid robot (e-puck) that has limited feedback mechanisms. The experiment was designed as a collaborative navigation task in which a human participant guides the robot using free hand gestures to follow a path projected on the ground that cannot be seen by the robot as accurately as possible. At different points of the path, virtual barriers exist that can be sensed by the robot but cannot be seen by the human operator. To succeed in completing the task, the human and the robot need to communicate their knowledge of the path and barriers. The communication channel used by the human was gesture but the robot had two possible communication channels: verbal feedback and nonverbal feedback. Nonverbal feedback was achieved by having the robot slow down when approaching a barrier and repeat forward-backward motions when the barrier lifts.

Gaze control

Gaze is one of the most important nonverbal behaviors in human-human interactions (Argyle, 2001). It was shown to be of comparable importance for natural HRI (Imai, Kanda, Ono, Ishiguro, & Mase, 2002; Sidner, Kidd, Lee & Leash, 2004; Yamazaki et al., 2008). Using appropriate gaze behavior increases people's engagement when interacting with a humanoid (Sidner et al., 2004) and when that behavior is contingent with the human partner it generates a stronger feeling of being looked at. Gaze-cuing behaviors were shown to increase as a result of the belief that an observed agent is an intentional entity and did not matter whether it had a face of a human or a robot (Wiese, Wykowska, Zwickel, & Müller, 2012).

These studies suggest that natural gaze behavior is an important asset for naturally interacting robots. Based on studies of human-human interactions and assumptions about humans' tendency to anthropomorphize robots, appropriate gaze behavior can be hard coded in the robot's behavioral repertoire. Thomaz and Breazeal (2006) used a modified reinforcement learning algorithm augmented with gaze and attention behaviors to improve the performance of standard reinforcement learning. More general gazing behaviors can be programmed in robots as well. Yonezawa, Yamazoe, Utsumi, and Abe (2007) utilized gaze tracking to achieve mutual gaze (looking at the interaction partner) and mutual attention (sharing gaze target with interaction partner) and showed that this simple manipulation increased the positivity of partner feelings toward the robot. Mohammad and Nishida (2010a) used a reactive robotic architecture called EICA and four simple motor plans activated and deactivated through three higher level interaction processes to achieve humanlike gaze behavior on a humanoid robot. The interaction processes implemented an approach-avoidance mechanism for gaze control inspired from spatial behaviors of humans during face-to-face situations and augmented with a third process to focus the robot's attention on the most salient feature of the environment as determined by a gaze-following based map. This design generated mutual attention and mutual gaze behaviors in similar proportions to their natural occurrences in human-human interactions even though these proportions were not programmed into the robot.

Spatial behavior and proxemics

Early work on robot navigation considered humans as a form of obstacle to be avoided for the safety of both the robot and the human. Recently, a large body of research started to grow that studied social aspects of robot's spatial behavior in the proximity of humans with the explicit goal of making robot move in a socially acceptable manner when interacting with people. This was employed in motion panning (Sisbot, Marin-Urias, Alami, & Simeon, 2007), socially acceptable human avoidance (Yoda & Shiota, 1996), and human-aware goal-directed navigation (Feil-Seifer & Matarić, 2011). Dondrup, Bellotto, and Hanheide (2014) employed a probabilistic representation to encode the joint spatial behavior of robots and humans and showed the appropriateness of this representation in producing socially recognizable spatial behaviors like passing-by situations, and avoidance in corridors. This work employed an augmented version of the qualitative trajectory calculus, which is a formalism used to represent relative motions of two points in space in qualitative terms. The use of QTC reduced the complexity of the problem by focusing on the qualitative terms relevant to the situation (Dondrup, Hanheide, & Bellotto, 2014).

Human-robot interaction situations are not always dyadic and some research focused on the effect of having more than two partners on spatial behavior. For example, Vázquez, Steinfeld, Hudson, and Forlizzi (2014) designed a study in which a robot (Chester) resembling a chiffonier either interacted with a group of children alone or accompanied by a sidekick robot resembling a lamb (Blink), which also interacted with the children. The interaction was designed in 11 stages starting with acknowledgement and greetings and ending with a Goodbye. Analysis of the children's spatial behavior around the robot showed that it can be modeled by a mixture of three Gaussians corresponding roughly to the intimate space at the distances 0.1 to 1.1 m from the robot, social space in which most face-to-face interactions occur at distances between 1.1 and 3.3 m and the public space at distances over 3.3 m. These three spaces corresponded roughly to Hall's intimate and personal spaces (0.15 to 1.2 m), social space (1.2 to 3.7 m), and public space (beyond 3.6 m). This similarity between the spatial distribution of children in relation to the robot and to human partners supports the argument at the root of natural HRI research that humans tend to anthropomorphize robots. It was also shown that children tended to face the robot with 99.7% of the time being within the front half-circle of the robot following a normal distribution with the mean at around 90 degrees (i.e. facing the robot directly). The existence of the sidekick did not change the spatial behavior of the children but elicited more attention to the robot's speech, which suggested a higher level of engagement.

Interaction, dyadic or not, happen within a context; and naturalness of robot behavior is expected to depend on the specific context. For example, when the robot is expected to physically pass objects to a human, it should interact within the intimate or personal space instead of the social space used in other cases. Koay et al. (2007) studied user preferences for the direction from which a robot should approach them in a can-passing task, the distance at which interaction is to start, the distance from which the robot should start the object passing motion and the specific trajectory by which this motion is to be executed. The results show that a majority of the participants prefer the robot to approach from the front and hand them a can of soft drink in the front sector of their personal zone. The most influential factor in determining the most appropriate approach position was the hand positioning during handing over the can. Legibility and perception of risk seem to be the factors that decide how participants choose their preferred robot approach coordination between the arm and the base for handing over the can.

Learning Interaction Protocols

Most of the systems considered in this chapter so far are designed using the standard engineering approach in HRI, in which features and rules governing human behavior in some interaction context are studied using either human-human interaction experiments, Wizard of Oz (WOZ) HRI sessions, or analysis of research results in interaction studies. These features and rules are then used to model *natural* behavior of people in that context and the models are then implemented into the control software of the robot. Even though human likeness is not always the same as natural robotic behavior, in most cases it is close enough. The robot is then evaluated in real-world HRI sessions and results of human evaluation of its behavior are used to tune the model.

This approach for natural HRI has the advantage of generating clear models that support the plausibility of robot's behavior based on the assumption of approximate equivalence between human likeness and naturalness for humanoids.

Another approach for generating equivalent models is to use machine-learning techniques for learning the appropriate behavior of the robot from corpora of human-human and human-robot interaction databases. This approach has the advantage of requiring much less manual tuning allowing the evaluation of the system to occur much faster than is usually the case with the standard engineering approach. This, in turn, has the potential of allowing more iterations in which the behavior of the robot is tuned, possibly using similar machine-learning techniques.

This section provides an overview of some machine-learning-based natural HRI systems. An important concept for these systems is the *interaction protocol*, which encapsulates the rules of interaction within some context. Interaction protocols can be represented using probabilistic models (Mohammad & Nishida, 2010b), coupled dynamical systems, or parallel processing with no predefined formal representation (Mohammad & Nishida, 2009). They can be flat representing the interaction at a single level of abstraction or hierarchical with shallow (Mohammad & Nishida, 2010c) or deep structure (Mohammad & Nishida, 2010b).

As an example of this approach, we consider the embodied interactive control architecture proposed by Mohammad et al. (2010). The architecture is built as a hierarchy of levels of specifications where each level provides its own control mechanisms and processes. The lowest level of specification is a distributed control architecture allowing low-level control processes (called intentions) to provide a shallow pathway between sensing and actuation. These processes can interact through different kinds of data channels providing the basic building blocks of robot software. This architecture is similar to the ROS architecture in most aspects and provides no specific support of services for HRI specific applications. The second level of specification (called LiEICA) is designed specifically to learn natural interaction protocols in the form of a hierarchy of interacting processes. The lowest level of these processes are called basic interactive acts and they are learned directly from the interaction corpus used for training the system to represent repeated patterns of behaviors in the interaction. Things like turning the head to face the interaction partner (e.g. mutual gaze), mutual attention to objects, simple object directed actions like grasping are encoded at this level. Higher levels of control are called interaction control layers and consist again of sets of processes that collaborate to model the interaction protocol at increasingly higher levels of abstraction.

Learning of interaction protocols is achieved in this system in three stages: During the first stage, basic interactive acts are learned by applying a motif discovery algorithm to the training data to segment repeating patterns of motion in the partners' behaviors during the interaction. Each such pattern is modeled either by a probabilistic model or a dynamic system. A reverse process is also learned to detect each of these learned models in continuous streams of sensor time series.

Having learned the basic interactive acts, the second stage allows the robot to learn a plan for activating them at increasingly higher levels of abstraction using a hierarchy of coupled dynamical systems.

After the second stage is completed, the robot can then engage in actual interactions of similar type with human partners and use the differences between their behavior and the prediction it has for that behavior based on the learned dynamical hierarchy to adjust the interaction protocol (Mohammad & Nishida, 2010a).

Another approach for a learning interaction protocol focuses on verbal instead of only nonverbal behaviors. The scenario used is a shopping scenario in which the robot behaves as the shopkeeper. Data was collected using 16 Kinect 1 sensors and participants had to touch a screen at the beginning and ending of their speech. 178 interactions were recorded including 1197 customer utterances and 1233 shopkeeper utterances.

The basic approach was similar in spirit to the EICA approach discussed earlier but different in the implementation and the algorithms used. Firstly, basic units of behavior were segmented from the data (abstracted). That included motion trajectories, spatial formations, speech vectorizations and robot speech actions. A learning algorithm was then utilized to allow the robot to predict the appropriate utterance or behavior based on the training data.

Human-Robot Interaction Corpuses

Records of HRIs can help advance the research in natural HRI for several reasons. First, they provide a standardized set of scenarios that can be used to compare different algorithms. Second, they provide training data for methods based on machine learning (e.g. the interaction protocol learning approach discussed earlier). Unfortunately, it is not easy to collect or utilize these recordings because of the high costs involved in collecting them. Moreover, their use for comparing different approaches to natural

HRI is limited by the fact that the final evaluation of such systems must rely on online HRI sessions that are not easily modeled after the recordings in a specific corpus of interactions. A few such datasets are available for researchers including the Vernissage Corpus (Jayagopi et al., 2013), HuHRIC (Bastianelli, 2014), and H^3R.

The HuHRIC corpus focuses only on verbal dialogue between a robot and a human. It provides audio files along with their transcripts referring to commands given to the robot in a home environment (Bastianelli, 2014).

The Vernissage corpus uses a NAO robot situated in a room and providing multiple humans with information about paintings on the wall. The data recorded included the videos from the NAO's cameras and three external cameras, VICON motion capture data of the humans involved, close-talk mics for the participants, the NAO's mic, and motion states. The corpus is also annotated with nonverbal cues related to the interaction. Speech transcripts and ground truth data including head poses, visual foci of attention of participants, nodding, utterances, addresses, and 3D locations are also provided (Jayagopi et al., 2013).

The H^3R corpus uses a Robovie II robot and provides human-human interactions as well as human-robot interactions. The corpus contains records of 66 sessions; 44 of these involve human-human interactions (half natural and half unnatural) in which one participant is explaining to another the assembly/disassembly of either a chair or a bicycle. The other 22 sessions record similar explanations by a participant to a robot that provided basic mutual gaze and mutual attention controls. Synchronized audio, video, and motion data (using the PhaseSpace sensor) are provided along with measurements of skin conductance, respiration patterns, and pulse signals from all participants. This data can be used to train interaction protocol learners and provide a baseline for comparing the behavior of different listener robots to both the base robotic controller used in the corpus as well as to human behaviors in the same situation.

Conclusion

This chapter discussed the concept of natural human-robot interaction, which was divided into naturally controllable robots and natural-behavior robots. The former are robots that can be controlled using natural means of communication. Both verbal and multimodal natural control were discussed, with examples from recent literature. The focus of the chapter was on naturally behaving robots. These are robots the *behave* naturally in some sense of the word. Different possible definitions of natural robotic behavior were discussed with a focus on nonverbal behaviors. As examples of the kinds of behaviors exhibited by these robots, we discussed gaze control and spatial behaviors.

Naturally behaving robots are traditionally developed by discovering patterns of behavior in human-human interactions and hard-coding them into the robot (what we called the engineering approach to HRI). A promising approach that is starting to gain momentum uses machine-learning techniques to have the robots learn natural interactive protocols from watching human-human interactions. To enable this technique, there is a need to have both natural and unnatural human-human and human-robot interaction recordings. Three datasets of such interaction recordings were discussed, providing both verbal and nonverbal data.

References

Argyle, M. (2001). *Bodily communication*. London: Routledge.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*(1), 37–46.

Bartneck, C., Kanda, T., & Norihiro Hagita, H. I. (2009). My robotic doppelganger: A critical look at the uncanny valley. In *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN '09)* (pp. 269–276). doi:10.1109/ROMAN.2009.5326351

Bastianelli, E., Castellucci, G., Croce, D., Iocchi, L., Basili, R., & Nardi, D. (2014). HuRIC: A human robot interaction corpus. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)* (pp. 4519–4526). Reykjavik: European Language Resources Association.

Bennewitz, M., Axenbeck, T., Behnke, S., & Burgard, W. (2008). Robust recognition of complex gestures for natural human-robot interaction. In *Proceedings of the Workshop on Interactive Robot Learning at Robotics: Science and Systems Conference (RSS)*. doi:10.1.1.143.4765

Bisio, A., Sciutti, A., Nori, F., Metta, G., Fadiga, L., Sandini, G., & Pozzo, T. (2014). Motor contagion during human-human and human-robot interaction. *PloS one*, *9*(8), e106172.

Burgard, W., Cremers, A. B., Fox, D., Hähnel, D., Lakemeyer, G., Schulz, D.,…Thrun, S. (1998). The interactive museum tour-guide robot. In *Proceedings of the 15th National/10th Conference on Artificial Intelligence/Innovative Applications of Artificial Intelligence* (pp. 11–18). Menlo Park, CA: American Association for Artificial Intelligence, Menlo Park, CA.

Chai, J. Y., She, L., Fang, R., Ottarson, S., Littley, C., Liu, C., & Hanson, K. (2014). Collaborative effort towards common ground in situated human-robot dialogue. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 33–40). doi:10.1145/2559636.2559677

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: the perception–behavior link and social interaction. *Journal of Personality and Social Psychology*, *76*(6), 893–910. doi:10.1037/0022-3514.76.6.893

Cicirelli, G., Attolico, C., Guaragnella, C., & D'Orazio, T. (2015). A kinect-based gesture recognition approach for a natural human robot interface. *International Journal of Advanced Robotic Systems*, *12*(3), 22.

Clark, H.H., (1996). *Using language*. Cambridge: Cambridge University Press.

Dondrup, C., Bellotto, N., & Hanheide, M. (2014). Social distance augmented qualitative trajectory calculus for human-robot spatial interaction. In *Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication (2014 RO-MAN)* (pp. 519–524). doi:10.1109/ROMAN.2014.6926305

Dondrup, C., Hanheide, M., & Bellotto, N. (2014, March 24–26). *A probabilistic model of human-robot spatial interaction using a qualitative trajectory calculus*. Paper presented at the Qualitative Representations for Robots Spring Symposium, Stanford University, CA.

Farroni, T., Csibra, G., Simion, F., & Johnson, M. H. (2002). Eye contact detection in humans from birth. *Proceedings of the National Academy of Sciences*, *99*(14), 9602–9605. doi:10.1073/pnas.152159999

Feil-Seifer, D., & Matarić, M. (2011). People-aware navigation for goal-oriented behavior involving a human partner. In *2011 IEEE International Conference on Development and Learning (ICDL)* (pp. 1–6). doi:10.1109/DEVLRN.2011.6037331

Gorostiza, J. F., Barber, R., Khamis, A. M., Malfaz, M., Pacheco, R., Rivas, R.,…Salichs, M. A. (2006). Multimodal human-robot interaction framework for a personal robot. In *Proceedings of the Fifteenth IEEE International Symposium on Robot and Human Interactive Communication, 2006 (ROMAN2006)* (pp. 39–44). doi:10.1109/ROMAN.2006.314392

Imai, M., Kanda, T., Ono, T., Ishiguro, H., & Mase, K. (2002). Robot mediated round table: Analysis of the effect of robot's gaze. In *Proceedings of the 11th IEEE International Workshop on Robot and Human Interactive Communication* (pp. 411–416). doi:10.1109/ROMAN.2002.1045657

Jayagopi, D. B., Sheiki, S., Klotz, D., Wienke, J., Odobez, J. M., Wrede, S.,…Gatica-Perez, D. (2013). The vernissage corpus: A conversational human-robot-interaction dataset. In *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 149–150). doi:10.1109/HRI.2013.6483545

Koay, K. L., Sisbot, E. A., Syrdal, D. S., Walters, M. L., Dautenhahn, K., & Alami, R. (2007). Exploratory study of a robot approaching a person in the context of handing over an object. Presentation at the *AAAI Spring Symposium: Multidisciplinary Collaboration for Socially Assistive Robotics* (pp. 18–24). Stanford, CA: Association for Advancement of Artificial Intelligence (AAAI).

Kollar, T., Vedantham, A., Sobel, C., Chang, C., Perera, V., & Veloso, M. (2012). A multimodal approach for natural human-robot interaction. In *Proceedings of the International Conference on Social Robotics* (pp. 458–467). doi:10.1007/978-3-642-34103-8_46

Lee, C., Lesh, N., Sidner, C. L., Morency, L. P., Kapoor, A., & Darrell, T. (2004). Nodding in conversations with a robot. In *CHI'04 Extended Abstracts on Human Factors in Computing Systems* (pp. 785–786). doi:10.1145/985921.985935

Matsui, T., Asoh, H., Fry, J., Motomura, Y., Asano, F., Kurita, T.,…Otsu, N. (1999). Integrated natural spoken dialogue system of Jijo-2 mobile robot for office services. In *Proceedings of the Sixteenth National Conference on Artificial Intelligence and the 11th Innovative Applications of Artificial Intelligence Conference: Innovative Applications of Artificial Intelligence* (pp. 621–627). Menlo Park, CA: American Association for Artificial Intelligence.

Mohammad, Y., & Nishida, T. (2007). Talkback: Feedback from a miniature robot. In *AI 2007: Advances in Artificial Intelligence* (pp. 357–366). Berlin: Springer.

Mohammad, Y., & Nishida, T. (2009). Towards combining autonomy with interaction for social robots. *AI and Society Journal*, *24*(1), 35–49.

Mohammad, Y., & Nishida, T. (2010a). Controlling gaze with an embodied interactive control architecture. *Applied Intelligence*, *32*(2), 148–163.

Mohammad Y., & Nishida T., (2010b). Learning interaction protocols using Augmented Baysian Networks applied to guided navigation. *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 4119–4126). doi:10.1109/IROS.2010.5651719

Mohammad Y., & Nishida T. (2010c). Learning spontaneous nonverbal behavior using a three layers hierarchy. *Tenth WSEAS International Conference on Applied Computer Science* (pp 430–434). Stevens Point, WI: World Scientific and Engineering Academy and Society.

Mohammad, Y., Nishida, T., & Okada, S. (2010, June). *Autonomous development of gaze control for natural human-robot interaction*. Presentation at the IUI 2010 Workshop on Eye Gaze in Intelligent Human Machine Interaction (IUI'10), Santorini.

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley. *IEEE Robotics and Automation Magazine*, *19*(2), 98–100.

Perzanowski, D., Schultz, A. C., Adams, W., Marsh, E., & Bugajska, M. (2001). Building a multimodal human-robot interface. *IEEE intelligent systems*, *16*(1), 16–21. doi:10.1109/MIS.2001.1183338

Rautaray, S. S., & Agrawal, A. (2015). Vision based hand gesture recognition for human computer interaction: A survey. *Artificial Intelligence Review*, *43*(1), 1–54.

Riek, L. D., Paul, P. C., & Robinson, P. (2010). When my robot smiles at me: Enabling human-robot rapport via real-time head gesture mimicry. *Journal on Multimodal User Interfaces*, *3*(1–2), 99–108.

Sidner, C. L., Kidd, C. D., Lee, C., & Leash, N. (2004). Where to look: A study of human-robot engagement. In *Proceedings of the Ninth International Conference on Intelligent User Interfaces (IUI '04)* (pp. 78–84). New York, NY: ACM.

Sisbot, E. A., Marin-Urias, L. F., Alami, R., & Simeon, T. (2007). A human aware mobile robot motion planner. *IEEE Transactions on Robotics*, *23*(5), 874–883.

Spiliotopoulos, D., Androutsopoulos, I., & Spyropoulos, C. D. (2001, June). *Human-robot interaction based on spoken natural language dialogue*. Paper presented at the European Workshop on Service and Humanoid Robots, Santorini, Greece.

Stiefelhagen, R., Fugen, C., Gieselmann, R., Holzapfel, H., Nickel, K., & Waibel, A. (2004). Natural human-robot interaction using speech, head pose and gestures. In *IEEE/RSJ International Conference on Intelligent Robots and Systems, 2004 (IROS 2004)* (vol. 3, pp. 2422–2427). doi:10.1109/IROS.2004.1389771

Swaminathan, R., & Sridharan, M. (2011). *Towards natural human-robot interaction using multimodal cues* (Technical Report). Lubbock, TX: Department of Computer Science, Texas Tech University.

Thomaz, A. L., & Breazeal, C. (2006, May). *Transparency and socially guided machine learning*. Paper presented at the Fifth International Conference on Developmental Learning (ICDL), Bloomington, IN.

Vallei, A. (2008). The design of natural interaction. *Multimedia Tools and Applications*, *38*(3), 295–305. doi:10.1007/s11042-007-0190-z

Vázquez, M., Steinfeld, A., Hudson, S. E., & Forlizzi, J. (2014). Spatial and other social engagement cues in a child-robot interaction: Effects of a sidekick. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 391–398). New York, NY: ACM.

Waldherr, S., Romero, R., & Thrun, S. (2000). A gesture based interface for human-robot interaction. *Autonomous Robots*, *9*(2), 151–173.

Wiese, E., Wykowska, A., Zwickel, J., & Müller, H. J. (2012). I see what you mean: how attentional selection is shaped by ascribing intentions to others. *PloS One*, *7*(9), e45391.

Yamazaki, A., Yamazaki, K., Kuno, Y., Burdelski, M., Kawashima, M., & Kuzuoka, H. (2008). Precision timing in human-robot interaction: coordination of head movement and utterance. In *Proceedings of the 26th Annual SIGCHI Conference on Human Factors in Computing Systems (CHI '08)* (pp. 131–140). New York, NY: ACM.

Yoda, M., & Shiota, Y. (1996). Analysis of human avoidance motion for application to robot. In *Fifth IEEE International Workshop on Robot and Human Communication* (pp. 65–70). doi:10.1109/ROMAN.1996.568749

Yonezawa, T., Yamazoe, H., Utsumi, A., & Abe, S. (2007). Gaze-communicative behavior of stuffed-toy robot with joint attention and eye contact based on ambient gaze-tracking. In *Proceedings of the 9th International Conference on Multimodal Interfaces (ICMI '07)* (pp. 140–145). New York, NY: ACM.

29

A Quick Look at Game Engagement Theories

Chaklam Silpasuwanchai and Xiangshi Ren

Background

Understanding game engagement is crucial to the success of digital games and emerging trends of gamification and serious games. This chapter provides a quick reference resource containing the commonly discussed game engagement theories and concepts. The primary readership is likely to be non-game researchers / practitioners who are interested in understanding digital games at the theoretical level. This chapter will allow them to overview the existing theories / concepts quickly. Likewise, game designers and researchers may find this overview useful.

Given that there are quite a number of theories, this chapter can only briefly discuss each theory / concept. It assumes that readers can carry out further research and apply specific theories to their own work. It is also important to note that this chapter will not cover game design heuristics, game research methodologies, or other subareas of games, including serious games and gamification, as these have already been covered extensively in past reviews of, for example, heuristics (Desurvire, Caplan, & Toth, 2004; Federoff, 2002), definitions and methodologies (Boyle, Connolly, Hainey, & Boyle, 2012; Mekler, Bopp, Tuch, & Opwis, 2014), gamification (Hamari, Koivisto, & Sarsa, 2014), serious games (Bellotti, Kapralos, Lee, Moreno-Ger, & Berta, 2013), and health games (Kato, 2010). Instead, this manuscript will focus on understanding game engagement mainly from the theoretical point of view.

Methodology

There are many theories scattered across different disciplines. Each theory has its distinct roots, so we contend that the synthesis of these theories can provide a more comprehensive understanding of game engagement.

This chapter presents the results of a systematic review. The methodology consisted of four steps: (a) identifying relevant databases and collections; (b) identifying relevant search terms / keywords; (c) specifying selection criteria, and (d) performing coding analysis.

The Wiley Handbook of Human Computer Interaction, Volume 2, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

Databases and data collection

We searched databases that are commonly used for publications in digital games, including those identified as relevant to information technology, psychology, and social sciences: ACM Digital Library, DiGRA, Web of Science, Scopus, IEEExplore, Science Direct and EBSCO. We restricted our search from 2000 to 2014, given the surge of interest in game studies since 2000 (Boyle et al., 2012) as seen in the establishment of game conferences such as ACE, Future Play, DiGRA, or SIGGRAPH Sandbox since the mid-2000s (some of them recently merged into the new CHI PLAY conference series).

To prevent possible loss of important theories before 2000 or other sources of information, we used the snowballing method (Jalali & Wohlin, 2012), where commonly cited papers shared across relevant documents were also collected. Thus we were able to retain important work, such as Csikszentmihaly's flow theory (Csikszentmihalyi, 1990) and Caillois's (Caillois, 1961) and Huizinga's play theory (Huizinga, 1950).

Search terms

We derived some search terms from previous reviews)Boyle et al., 2012; Mekler et al., 2014). We also considered that engagement is often associated with many other terms such as "fun" and "motivation." Our resulting search terms include: "Games," "Engagement," "Enjoyment," "Fun," "Motivation," "Attention," "Emotion," and "Affect." Through snowballing, we further added the following search terms: "Play," "Immersion," "Presence," "Embodiment," "Social," and "Relationships."

Selection criteria

Our search terms resulted in many papers being retrieved from the databases. To keep things manageable, we narrowed the scope for inclusion of papers using the following three screening criteria. (a) The documents had to provide a theoretical contribution to game engagement. (b) We intended to focus only general games, thus other subcategories of games were omitted. (c) Finally, the documents had to be original, peer-reviewed papers written in the English language. Using these three limiting conditions 74 documents were included for the current review.

Coding

Each document was initially coded. The initial codes include the theory/concept, published year, related factors and field of study.

We further conducted a *thematic analysis* (Braun & Clarke, 2006) on our collection of theories and studies. The coding was conducted in entirety by the authors. To confirm interrater reliability, an independent rater performed the same coding using 50% of the documents. Interrater reliability was found to be Kappa = 0.926 ($p < 0.01$), which is considered high. Slight disagreement happens in broad theories such as play theory. We expected and considered that certain theory may be difficult to classify given their broad coverage. We used our best judgement to classify these broad theories.

Table 29.1 Classification framework.

Themes	*Description*
Needs satisfaction	Players' nature and fundamental nature and psychological needs are in focus here. How do games satisfy players' psychological needs? What are players' psychological needs?
Emotions	Players' emotional responses and feelings are affective factors. What is the impact of games on emotions? How do positive and negative valence impact players' engagement?
Cognition	Players' thought processes, perception, embodiment, and visual attention affect game engagement. What is the effect of games on players' cognition (e.g., loss of awareness)? How does one reach those states?
Relationships	The impact of social influences on players' engagement is in focus here. What governs a positive and negative social experience? How do social experiences impact players' engagement as a whole?
Aesthetics	Game aesthetics impact game engagement. How do games visually and audibly affect players? How realism influences player experience?

The analysis revealed five high-level themes—(a) needs satisfaction, (b) emotion / affect, (c) cognition, (d) relationships, and (e) aesthetics. The final classification framework is described in Table 29.1 and the final inclusion of theories are listed in Table 29.2.

Theme 1: Needs Satisfaction

This theme facilitates understanding of game engagement by discussing the characteristics of human nature.

Self-determination theory

In self-determination (SD) theory (Deci & Ryan, 2000), Deci and Ryan classified motivation into two categories: *intrinsic* motivation, driven by inner needs, and *extrinsic* motivation, driven by external factors, such as rewards or threats. They described intrinsic motivation as drives, which all human beings will strive to meet: their inner drives for *autonomy*, *competence*, and *relatedness*. The feeling of independence, being in control of things, feeling the "origins" of their own actions, and making their own choices are primary inner drives for *autonomy*. *Competence* is the drive to fulfill one's need of feeling in control through skill mastery. Lastly, *relatedness* is the need to connect, to interact, to be accepted, and to be understood. While human beings prefer to be independent and competent, they are also motivated by acknowledgement from others regarding their independence and competence.

Ryan et al. (2006) argued that SDT theory can be used to explain the underlying motivational structure of video games, because the theory focuses on human basic needs, which are found across games and player types. Through a series of studies, Ryan et al. found that the SD theory—cf. the player experience of need satisfaction (PENS) model—can be used to predict enjoyment and long-term engagement.

Table 29.2 Overview of theories.

Theory/ concept	*Author(s) year*	*Field/ basis*	*Related evidence in games*
Needs satisfaction			
SD theory	Deci and Ryan (2000)	Psychology	Ryan, Rigby, and Przybylski (2006) found that autonomy, competence and relatedness predict enjoyment and long-term engagement
U&G theory	Katz and Blumler (1974)	Media Psychology	Sherry, Lucas, Greenberg, and Lachlan (2006) analyzed the uses and gratifications of games
Emotion			
Emotion theory	Various	PX (Player Experience)	Lazzaro (2004) identified emotions in games
Play theory	Caillois (1961); Huizinga (1950)	Culture	Caillois (1961) observed and classified play in traditional games such as poker and board games
Mood management theory	Zillmann and Bryant (1985)	Media Psychology	Reinecke et al. (2012) found that games can repair mood
Affective disposition theory	Raney (2003)	Media Psychology	Klimmt, Rizzo, Vorderer, Koch, & Fischer (2009) studied disposition (suspense) in games
Cognition			
Flow theory	Csikszentmihalyi (1990)	Positive Psychology	Sweetser and Wyeth (2005) developed GameFlow model
Immersion	Various contributors	PX	Jennett et al. (2008) studied immersion in games and developed immersion questionnaire
Presence	Various	Media Psychology; PX	Slater, Usoh, and Steed (1994) studied depth of presence in virtual environment
Embodiment theory	Various	Media Psychology; PX; Whole-body interaction	Bianchi-Berthouze (2013) explored how body movement affects player engagement
Goals	Various	PX	Andersen et al. (2011) studied how secondary goal can degrade game engagement
Relationships			
FIRO theory	Schutz (1958)	Social Psychology	Lucas and Sherry (2004) studied FIRO theory in games
Social facilitation theory	Zajonc (1965)	Social Psychology	De Kort, Ijsselsteijn, and Poels (2007) studied social presence and developed SPGQ
Social comparison theory	Festinger (1954)	Social Psychology	Ryan et al. (2006) show the impact of game experience on self-esteem
Aesthetics			
Audiovisual	Various	PX	Nacke, Grimshaw, and Lindley (2010) found the absence of sound affects game engagement
Realism and fidelity	Various	Media Psychology; Simulation; PX	Ivory and Kalyanaraman (2007) found the increased realism can enhance sense of presence in games.

Similar results were presented in related literature (Tamborini, Bowman, Eden, Grizzard, & Organ, 2010).

Uses and gratification theory

Uses and gratifications (U&G) theory (Katz & Blumler, 1974; Lucas & Sherry, 2004) describes how people engage with media (e.g., games) to satisfy their specific needs (e.g., enhancing knowledge, escape, relaxation). The theory further describes how every individual has different needs based on their past experiences, interests, and motives. For example, some people might engage with a game for relaxation, whereas others might play games to fulfill their need to feel competent. Finally, U&G theory assumes that players are active audiences who have control over what games they would play, suggesting that game engagement is voluntary and selective. Using U&G as a theoretical foundation, researchers have identified several uses of video games, e.g., to relax / escape / kill time / avoid doing other things (Phillips, Rolls, Rouse, & Griffiths, 1995; Sherry et al., 2006), to compete (Sherry et al., 2006; Yee, 2005), to achieve (Malone, 1980; Yee, 2005), to socialize (Sherry et al., 2006; Yee, 2005), to be aroused (Freeman, 2003; Lazzaro, 2004), to explore / discover / learn (Malone, 1980; Yee, 2005), and to fantasize (Malone, 1980; Sherry et al., 2006).

Theme 2: Emotion

This theme provides a means of understanding engagement through the lens of emotions.

Emotion theory

Researchers (Freeman, 2003; Lazzaro, 2004; Ravaja et al., 2004) have suggested that games are engaging because of the ability of games to evoke many different emotions. Ravaja et al. (2004) defined emotion as "biologically-based action dispositions that have an important role in the determination of behavior." Emotion contains three dimensions: (a) *subjective experience* (e.g., feeling happy), (b) *expressive behavior* (e.g., smiling), and (c) *physiological activation* (e.g., sympathetic arousal). Emotion (especially for physiological game assessment studies) is often described using a two-dimensional circumplex model (Russell, 1980) with two factors—valence (negative or positive emotions) and arousal (intensity of the emotions).

Ravaja et al. (2004) stated that games are successful because they are able to elicit a wide range of strong emotional responses, from fun and satisfaction to guilt and sadness. Lazzaro (2004) found over 30 different emotions that make games fun and enjoyable. Lazzaro synthesized these emotions info four different types of fun: hard fun (frustration), easy fun (wonder, curiosity), serious fun (excitement, relief) and people fun (recognition and connection). Hunicke, Leblanc, and Zubek (2004) identified eight types of fun: sensation, fantasy, narrative, challenge, fellowship, discovery, expression, and submission. More broadly, sensitivity theory (Reiss & Wiltz, 2004) defines fun as the satisfaction of 16 basic human desires (e.g., curiosity, power), with each of these desires linked to a particular emotion.

Enjoyment is a term that has been often associated with positive affect and identified as an affective outcome of a good gaming experience (Vorderer, Klimmt, & Ritterfield, 2004). Enjoyment is consistently regarded as key explanation for game engagement. Mekler et al. (2014) described enjoyment as the valence (affective aspect) of the player experience (fun, interest, pleasures). Blythe and Hassenzahl (2004) described enjoyment in the dimension of distractions.

Play theory

In Huzinga's play theory (Huizinga, 1950), he argued from a cultural perspective that "play" is essential to all human beings stating that "play is older than culture [...] all culture is an element of play." He further added that the most significant aspect of play is fun. Caillois (1961) described four fundamental types of play: Agon (competition), Alea (chance and uncertainties), Mimicry (role-playing), and Ilinx (changing state of mind and perception). He further described play along a dimension of interactive freedom—Ludus being a rule-based form of play and paidia being free-form improvisational play. Voluntary play or freedom of play (without third-party purpose) is a core aspect of games. One possible explanation of why humans enjoy the voluntary nature of play is that gamers are able to interact and express themselves more freely and emotionally, while not feeling controlled or monitored (Caillois, 1961).

Mood management theory

Based on the assumption that humans are pleasure-seeking beings, mood-management theory (Zillmann & Bryant, 1985) states that to maximize pleasures, humans instinctively tend to expose themselves to favorable environmental stimuli such that positive valence (pleasures) is maximized, whereas negative valence (pain) is minimized. Reinecke et al. (2012) studied mood management in games and found that mood management can result from mood repair through needs satisfaction. This theory is linked to selective exposure theory (Zillmann & Bryant, 1985), which states that humans possess tendencies to expose themselves to information that reinforces their previous beliefs or views while avoiding contradictory information. Nevertheless, this theory does not address why some players engage with negatively valenced activities, such as scary interactions in horror games, where fear and suspense are the primary emotions. One explanation was proposed by Klimmt (2003), which argued that players may engage with games that elicit negative feelings because they anticipate a resolution that will not only alleviate the negative feelings, but will result in feelings of euphoria and a great sense of achievement.

Affective disposition theory

Affective disposition theory (Raney, 2003) states that players make dispositional judgment of and emotional reactions to characters in the media/virtual world, which in turn affects their pleasures and enjoyment. For example, players have a tendency to share the sympathy and hope of the main character, while hoping for a negative outcome for the villain. The theory suggests that love (hero) and hate (villain) are two strong emotions that makes story enjoyable and engaging.

Disposition theory may provide an explanation for why "role playing" could be particularly enjoyable, as the player him/herself has the opportunity to facilitate negative outcomes for disliked characters, and to directly enjoy the victories of the "liked" protagonists. Researchers considered role playing as one of the most immersive type of game (Sweetser & Johnson, 2004; Yee, 2005). The interaction between "characters" and the "story" is critical for RPGs because it allows games to evoke a wide range of strong emotions, which make games engaging (Freeman, 2003; Sweetser & Johnson, 2004; Yee, 2005). In addition, the combination of a story with frustrations, dilemmas, decision making and multiple paths enables players to experience a deep level of emotional engagement and purpose (Freeman, 2003; Isbister, Flanagan, & Hash, 2010).

Theme 3: Cognition

This theme describes game engagement from the perspective of cognition.

Flow theory

Researchers (Cox, Cairns, Shah, & Carroll, 2012; Lucas & Sherry, 2004; Weibel & Wissmath, 2011) defined flow as the cognitive aspect of experience (involvement) with the task. Flow (Csikszentmihalyi, 1990) was defined as a mental state of being completely immersed, losing complete awareness even of bodily needs, with attention completely dedicated to a particular task at hand. He added that flow occurs when there is an optimal alignment between a user's skill level and the challenges posed by the task. There are seven additional elements needed to support that optimal alignment: clear goals, merging of action and awareness, concentration, autonomy, loss of consciousness, time distortion, and autotelic experience.

Past works (e.g., Piselli, Claypool, & Doyle, 2009) have supported flow theory by showing that the most satisfying and engaging moment for players is when there is an optimal alignment between the player's level of skill and the challenges provided by the game (e.g., barely victorious), while, when the challenge is too easy or too difficult (e.g., totally victorious) for players, the game becomes less engaging. Flow theory has been used extensively to explain the phenomenon of game engagement—for example, the GameFlow model (Sweetser & Wyeth, 2005). Nevertheless, Nakamura and Csikszentmihalyi (2002) found that enjoyment may occur independently of flow (i.e., flow describes an extreme gaming experience, which may not cover more casual experiences of enjoyment and lightweight absorption).

Immersion

Jennett et al. (2008) considered immersion as a result of good gaming experience. Immersion (Brown & Cairns, 2004; Cheng & Cairns, 2005; Jennett et al., 2008; Sanders & Cairns, 2010) comprises three main features: (a) temporal dissociation, (b) spatial dissociation and (c) merging of task and self. Jennett et al. (2008) argued that immersion is different from flow in the sense of extremity, i.e., since immersion is a less extreme version of flow, thus immersion can be used more effectively to describe a variety of player experiences (e.g., in casual gaming). Brown and Cairns (2004) defined immersion as the degree of involvement within gameplay, ranging from low (engagement) to moderate (engrossment) to high immersion (total immersion). Douglas and Hargadon (2000) viewed immersion as one of the primary sources of pleasures. Curran (2013) classified five types of immersion—general immersion, vicarious immersion (involved with the world and characters), action visceral immersion (involved with action and play), mental visceral immersion (involved with the tactics and strategies), and group immersion (involved with cooperative play).

Presence

Presence is closely related to immersion. Most commonly, researchers (Slater et al., 1994; Witmer & Singer, 1998) define presence as the sense of being there in the virtual environment without actually perceiving the existence of the medium. (Some refer presence to both task absorption and transportation into the environment. To avoid confusion, this chapter refers to presence specifically as spatial presence.) Researchers (Slater et al., 1994; Witmer & Singer, 1998) have found that the naturalness of the interactions and realism affect presence. Baños et al. (2004) found an association between emotion and presence—affective content (story) increases presence in a virtual simulation role-playing game. Jennett et al. (2008) argued that presence is only a small part of gaming experience; for example, one may experience immersion without presence in a puzzle game. On the other hand, one may experience presence without immersion, as when performing a boring task in a virtual simulation world. It can also be arguably said that presence is synonymous with spatial immersion (Weibel & Wissmath, 2011).

In terms of the impact of presence on enjoyment, Lombard and Ditton (1997) suggested that a high sense of presence leads to greater enjoyment. On the other hand, Weibel and Wissmath (2011) stated that the impact of presence depends on the types of games—presence is more important in vivid, realistic games like first-person shooters / role-playing games than in abstract puzzle / memory games, which require less realism.

Embodiment theory

Researchers (Bayliss, 2010; Benford & Bowers, 1995; Gee, 2008) described game experiences as an embodied phenomenon. Embodied cognition (Gee, 2008) refers to the state in which mind and body are connected and how they influence one another. It is specifically argued that bodily experiences can influence cognitions, unlike previous assumptions that envisaged the mind and body as separate entities. The concept of embodiment has often been used by researchers to describe the experience in full-body games (Bianchi-Berthouze, 2013), which have shown that body movement influences one's emotions and engagement. Embodiment also has been used to describe some role-playing games (Bayliss, 2010; Benford & Bowers, 1995; Gee, 2008), where the player's mind is influenced by the avatar's bodily experiences (i.e., the concept of embodiment can explain how players can become one with the avatar and feel deeply immersed). Embodiment illuminates our understanding that the mind, the body, and the environment (input device, outdoor, indoor) are all connected, which influences the player's engagement. This entirely suggested that cognition (embodied cognition) is not solely composed of the mind, but also influenced by bodily actions.

Goals

Clark (2007) argued that having goals is one of the primary motivating factors of games. Goals allow players to focus their attention fully onto one task (Clark, 2007). Habgood (2007) studied the correlation between extrinsic goals (e.g., rewards) in games and intrinsic goals (e.g., learning) and found that games are most powerful when the extrinsic goal of the game is aligned to the intrinsic goal. On the other hand, Andersen et al. (2011) studied secondary objectives in games. They found that secondary objectives such as collecting items or rewards that do not align with the primary goal of the game produced undesirable effects, such as boredom and frustration in players. Conversely, they found that these effects can be easily reversed by simply adjusting secondary objective items so that they align with the primary goal. When this is achieved, players become more engaged. In our view, users tend to be more engaged when they have one clear goal to focus on. By contrast, when there are many distracting goals (which do not have any strong or direct connection to the primary goal), users tend to become less engaged.

Theme 4: Relationships

This theme considers how social interaction can positively or negatively influence game engagement.

Fundamental interpersonal relationship orientation theory

Fundamental interpersonal relationship orientation (FIRO) Theory (Schutz, 1958) argues that all humans are governed by three social needs: inclusion, affection, and control. Inclusion refers to the need to belong to a social group and the need to interact with others. Affection refers the need to feel the sense of love and warmth in relationships. Control refers to the need to have influence/control over others' decisions/actions. Fundamental interpersonal relationship orientation theory also asserts that these orientations and priorities vary across people. Lucas and Sherry (2004) argued that these three factors can be exploited to enhance game engagement, for example by structuring gameplay around collaboration and competition to satisfy players' need for inclusion, affection, and control.

Social facilitation theory

Social facilitation theory (Zajonc, 1965) states that people have a tendency to perform differently in the presence of other people. Specifically, in the presence of other people, one would perform better in very familiar tasks but perform worse in less familiar tasks. Researchers mostly agreed that the changed performance is a result of awareness of possible evaluations from others, which can be readily observed in competition or collaboration scenarios in games. Nevertheless, the degree to which a person is influenced by social presence varies. De Kort et al. (2007) developed the Social Presence in Gaming Questionnaire (SPGQ) with social presence of others (e.g., playing with friends) as the primary source of motivation in gameplay.

Other relevant theories include social proof theory (Cialdini, 2008), which predicts that players are likely to engage in behaviors that others are also engaged in, while social identity theory (Tajfel & Turner, 1979) describes that humans share a sense of who they are based on their social groups (e.g., countries, gender, affiliations) as a process of self-image enhancement. Beenen et al. (2004) suggested that individuals are most socially motivated when their uniqueness and contribution is being acknowledged in a team environment.

Entirely, these theories suggest that humans are social in nature and that they seek the approval and avoid the disapproval. As a result, social factors such as social identity and status play important roles in enhancing game engagement.

Social comparison theory

Social comparison theory (Festinger, 1954) states that social experiences are driven by the need to better understand the self (accurate self-evaluations) as well as the need to improve one's self-esteem. This relationship between self-evaluations, comparison with others, and self-esteem implies that needs pertaining to competence and relatedness in self-determination may be associated. For example, in an online game environment, players, driven by the need to improve their self-esteem, may seek self-enhancement and verification from others about their skills level. If this observation is correct, it also implies that social experience may also partly driven by the need of competence. Thus social mechanisms, such as pushing high scores to the leader board, sharing trophies on public Web space, or even showing off their skills in public or with their friends may further promote sense of competence.

Theme 5: Aesthetics

This theme describes game engagement from the perspective of aesthetics—audio, visuals, and realism.

Audiovisual

Music and sound engage users by evoking and enhancing the intensity of emotions (Rossoff, Tzanetakis, & Gooch, 2010). Parker and Heerema (2007) described that sound creates a feeling of presence, reminding gamers that the game is still going on. Fast music may represent a lot of activity, and vice versa for slow music. They also suggest that sound affects emotions faster than visual display. Nacke et al. (2010) found significant correlations between audio and game engagement constructs.

In terms of visuals, LaViola and Litwiller (2011) found that players enjoyed playing using a 3D stereo display compared to a 2D display. Ermi and Mäyrä (2005) found that audiovisual capability and visual-motor links are fundamental in enabling a higher quality of gaming experience such as immersion. Takatalo, Häkkinen, Komulainen, Särkelä, and Nyman (2006) reported that screen size has no significant impact on engagement, although Banos et al. (2004) found otherwise. It appears that the importance of visual fidelity depends on the type of games (more important in role-playing / first person games). In terms of graphical aesthetics, Andersen, Liu, Snider, Szeto, and Popović (2011) found that gameplay variations affected play time three times as much as a variation in aesthetics. This finding suggests the supporting role of aesthetics on the overall gameplay.

Realism and fidelity

Realism is the extent to which a game resembles the real world. Realism is affected by the quality of aesthetics (visual and audible) in games, as well as the surrounding environment of players during gameplay. A similar term is fidelity, which Hays and Singer (1988) defined as the "degree of correspondence between simulation and real circumstances." Fidelity may cover the broader scope of realism to include physics and natural laws. Often, the more realistic the game, the more easily players feel a higher sense of presence, and more easily become immersed in the game, especially in vivid and realistic games (e.g., first-person shooting or simulation game) (Hays & Singer, 1988; Mcmahan, 2003; Slater et al., 1994; Witmer & Singer, 1998). Several studies about realism have been conducted, e.g., artificial gun vs. mouse (Kim, Biocca, & Jeong, 2011); large screen versus PC monitor (Baños et al., 2004); stereoscopic 3D games versus 2D games (Schild, LaViola, & Masuch, 2012). These studies indicate that realism increased level of presence; however, Weibel and Wissmath (2011) implied that realism plays a more important role in vivid, realistic games (first-person shooting / role-playing game) than in other puzzle / abstract games that require less realism.

Discussion

Needs satisfaction

From the review, we can better understand about game engagement. Game engagement has been often associated with needs satisfaction in SD theory and U&G theory. Needs satisfaction is considered by many researchers to be the key explanation of game engagement, where various psychological needs have been identified and mapped. SD theory describes explicit, high-level needs including autonomy, competence, and relatedness, while U&G covers broader range of needs including implicit needs such as relaxation and pleasure. Game engagement has also been identified as a selective and voluntary process in U&G theory, suggesting that game engagement varies across different persons.

There is strong evidence (Ryan et al., 2006; Tamborini et al., 2010; Yee, 2005) that games satisfy the need for autonomy, competence, and relatedness. There are also evidence that games satisfy other needs such as escape and relaxation (Sherry et al., 2006). Researchers found that needs satisfaction predicts long-term engagement (Ryan et al., 2006; Tamborini et al., 2010). Among all the needs, challenge (competence) is consistently rated as the key factors for engagement in SD theory (Ryan et al., 2006) and in U&G theory (Sherry et al., 2006).

Self-determination theory is often criticized for identifying narrow range of needs and U&G theory identified a broader range of needs, e.g., escape, relaxation. It seems that other needs will be identified in the future.

Uses and gratifications theory has also stated that needs vary in extent across people based on their past experiences, interests and motives but there has been lack of integrated understanding of how exactly these needs vary across person. This raises some interesting research questions whether needs are innate (always the same) with every human, or whether there are some needs that are affected by the existing environment and keep changing based on one's past experiences. This question may partly explain by the notion of attitudes (Nabi & Krcmar, 2004). Based on the technology acceptance model (TAM) (Davis, Bagozzi, & Warshaw, 1989) and the theory of planned behavior (TPB) (Ajzen, 1985), players' attitudes shaped player motives (e.g., to just relax, to win) in which player needs in turn determine the behaviors and engagement during gameplay. This process loops back to affect player attitudes. By viewing attitudes as individual characteristics we may better understand how engagement can also differ based on individual differences and external factors such as social norms.

Uses and gratifications theory views needs satisfaction as a selective and voluntary process. What intrigues us is why games may be chosen over other media or activities as other activities may also provide the same type of needs satisfaction. This leads to the recognition that features of the games themselves also contribute to game engagement such as cheap failure (Tocci, 2008), difficulty adjustment (Yun, Shastri, Pavlidis, & Deng, 2009), customization (Ducheneaut, Wen, Yee, & Wadley, 2009), tutorials (Andersen et al., 2012), rewards (Berkovsky, Coombe, Freyne, Bhandari, & Baghaei, 2010), choices (Isbister et al., 2010), and so forth. O'Brien and Toms (O'Brien & Toms, 2008) associated these features with "engagement attributes" including interactivity, perceived control, and novelty.

It has been stated that needs satisfaction may in part explain how game engage users (Deci & Ryan, 2000; Lucas & Sherry, 2004; Tamborini et al., 2010). Oliver and Raney (Oliver & Raney, 2011) used the word "eduaimonic" to link these needs pertaining to wellbeing, purpose, and meaningfulness. It can be argued that humans are purpose-seeking beings (Oliver & Raney, 2011). The evidence that players are purpose-seeking beings can be reflected from players' reported motives (Sherry et al., 2006; Yee, 2005) including the motives to win, to make progress, to interact with others, to explore, discover, and learn, and so forth, which are consistent with SD theory (Deci & Ryan, 2000) and U&G theory (Katz & Blumler, 1974).

Emotion

Game engagement has also been associated with emotions. Some have treated game engagement in the same fashion as enjoyment (positive affect) while others treated enjoyment as a key explanation for game engagement. Some also treated enjoyment as an affective component of game engagement. In any case, researchers found that engagement can occur in a negative-valenced gameplay (e.g., horror gameplay) suggesting that game engagement is associated with both positive and negative affect.

This theme offers the theoretical perspective that games are engaging because of their capability to evoke a wide range of strong emotions—both positive and negative. There is evidence that games elicit a strong, wide range of emotions (Ravaja et al., 2004).

Play theory, mood management theory, and affective disposition theory have provided theoretical explanations about why humans are attracted to pleasures and emotional arousal from different perspectives (e.g., culture, mood, disposition). They also provided some mappings to game features that contribute to strong emotions including uncertainties, difficult challenge, role-playing / story, competition, and dilemmas.

Enjoyment is a term associated with positive affect and is consistently rated as key explanation for game engagement (Mekler et al., 2014). However, engagement is also associated with negative affect (Jennett et al., 2008) suggesting that engagement may not occur due to enjoyment only but also negative arousals such as suspense, guilt, frustrations (Vorderer et al., 2004). Further research should investigate in greater detail how negative emotions impact game engagement and how games can be designed to elicit these emotions. It would also be beneficial to understand how emotion should be best designed—whether designers should design to elicit a wide range of *similar* emotions or to elicit a wide range of *different* emotions. Based on sensitivity theory, further research should also investigate how negative emotions relate with players' needs satisfaction, desires, values, and individual differences.

Past studies suggested a correlation between needs satisfaction and emotional arousal (Oliver & Raney, 2011; Reinecke et al., 2012; Tamborini et al., 2010). The feelings of need satisfaction is associated with user's experience of positive valence (Reinecke et al., 2012). It has also been suggested that high-level needs satisfaction and emotional arousal may compete with each other in certain times (Klimmt, 2003; Norman, 2003). For example, in negatively valenced games such as horror games, although the games may elicit negative emotions, these negative emotions may be hindered by the player's need of competence, to solve this challenge of fear, and once these challenges meet resolution, players feel a sense of accomplishment as well as the feeling of enjoyment at the same time (Klimmt, 2003). Perhaps this can partly explain how negative emotions engage users.

Cognition

Game engagement has also been discussed in the dimension of cognition, absorption, and distractions, namely the concepts of flow, immersion, and presence. It has been stated that when players are engaged, they can achieve flow, immersion or presence—a state where their awareness is dissociated spatially and temporally. Flow theory suggests that eight components are required to achieve the state of flow: clear goals, merging of action and self, concentration, autonomy, loss of consciousness, time distortion, autotelic experience, and alignment between challenge and skills.

Game engagement has been associated with spatial and temporal awareness, described by the concept of flow, immersion, and presence. While flow describes optimal experience, Jennett et al. (2008) argued that immersion is a more useful concept than flow as it can be used to explain a more casual gaming experience. Presence was often referred as teleportation to a virtual environment and may occur independently of immersion. Flow and immersion can be both seen as the motives of playing games (Yee, 2005), as well as a cognitive outcome of a good gaming experience (Piselli et al., 2009).

Nakamura and Csikszentmihalyi (2002) found that enjoyment may occur independently of flow. Mekler et al. (2014) stated that enjoyment is different from flow, i.e., enjoyment is a characteristic of flow, but enjoyment may occur independently of flow. One may view enjoyment as the affective aspect of the gaming experience (Vorderer et al., 2004), while absorption (flow, immersion) is the cognitive aspect (involvement) of the experience (Piselli et al., 2009). But it remains unclear which occurs first during an engagement process and how absorption and enjoyment relate, and how they sequentially contribute to engagement. Understanding this would help guide researchers studying the experimental design process to determine the relevant variables to measure and the statistical relationships to look for.

Embodiment theory has provided an interesting perspective on cognition. There is evidence that bodily interactions affect cognition (Bianchi-Berthouze, 2013). Our mind or cognition is affected by how we act on the environment, thus suggesting that game engagement is affected by bodily interactions with the environment, e.g., game controller, avatar, physical environment. Embodiment might also provide explanation regarding the difference in engagement between physical board/card games and their virtual counterparts. Designing game engagement will thus also need to consider the medium and environment of gameplay.

Relationships

At the social level, game engagement has been associated with relationships, as seen in many online games. Because players are motivated by the need to connect and to be approved by friends, social interaction impacts whether and how long a player will engage in a game. Relationships have also been closely related with the sense of feeling in control, self-esteem, and competence. This is clearly seen in the MMRPG phenomenon, in which player relationships form an important part of the motivation to play.

Relationships can be viewed as involving needs satisfaction in SD theory (relatedness). Relationships are also closely associated with the feeling of competence (Festinger, 1954), self-esteem (Ryan et al., 2006) and feeling in control (Schutz, 1958). This association was reflected in experiments where social presence of others affect one's engagement and performance (Zajonc, 1965). There is also evidence that humans seek acknowledgement from others for their uniqueness (Beenen et al., 2004) and competence (Ryan et al., 2006). Further research should include investigating how different types of social presence (e.g., physical friends, online friends) affect game engagement.

Yee (2005) identified socializing along with achievement and immersion as important motives for engaging games. In online games, socializing is identified as the key reason for playing (Ryan et al., 2006) suggesting that socializing may be more important in some games. Some common game features contributing to social engagement include teamwork, communication channels, competition, and leaderboards.

Although relationships can improve player engagement, it was implied that social features should be carefully designed so that feelings of alienation or reduced self-esteem should be avoided or minimized (Festinger, 1954). Care should also be taken not to design social features that disrupt immersion or flow as real people in social interactions may provide a link back to the real world (Sweetser & Wyeth, 2005).

Aesthetics

Game engagement has also been associated with our senses—how something looks and feels. These visceral features (audiovisuals, realism) affect our initial engagement and may disrupt our level of immersion and enjoyment when not correctly designed.

This theme concerns audiovisual level of experience. There are evidence that audio and visual cues facilitate game engagement (Nacke et al., 2010). When they are not carefully designed, they may cause disruption in absorption or lower enjoyment. Realism is seen by researchers as important in realistic games (e.g., Role playing) but not necessary in other types of games (e.g., puzzle game). It is also important for designer to be careful of the "uncanny valley" phenomenon (Mori, MacDorman, & Kageki, 2012), where the realism is extreme but lacks small cues, which makes the experience worse than a less realistic presentation. This often happens when the level of realism mismatch users' expectations.

Practical Guidelines

In design, one can consider how to design games based on these theories. There are four design considerations—autonomy (need satisfaction), competence (need satisfaction, cognition), relatedness (need satisfaction, relationship), and emotional arousals (emotion, aesthetics) (see Table 29.3).

Autonomy

For autonomy, players should feel that they personally orchestrate game outcomes, rather than that game designers predetermine how the game should unfold. This sense of autonomy can be enhanced by providing meaningful choices that lead to different game endings, a vast world to explore with different and surprising possibilities, different means to solve a challenge, or free space where players can experiment with their creativity. One example is the game of chess. In such games, there are thousands of possibilities regarding the development of the game, depending on the players' choices; players can experiment with various choices as the game unfolds.

Table 29.3 Practical guidelines from the theories.

Autonomy	*Competence*	*Relatedness*	*Emotional arousals*
Choices	Ability to customize	Collaboration	Story
Exploration	Feedback	Leaderboard	Aesthetics
Different possible	Cheap failure	Communication	Realism
solutions	Clear goal	Public space	Role playing
Not reward centered	Short- and long-term goal	Representations	Decisions
	Increasing difficulty		Personalization
	Tutorials		Risks, uncertainty and competition
	Awards and trophies		Time pressure

It is best not to overly control how users should use the gaming application, but provide the freedom for users to decide the best approach. For example, in a popular city-building game called SimCity, players are allowed to imagine freely how their ideal city could look and they are free to build it accordingly. Another example is a popular simulation game called Surgeon Simulator where the player, role playing as a surgeon, is given free space to experiment with different ways to perform a surgical procedure given a variety of tools and to look at the possible consequences. Game mechanisms that promote a sense of freedom and perceived control include:

- *Choices.* By enabling gamers to make decisions that affect the gaming outcomes, feelings of initiatives and control enhance game engagement. Many successful games such as the Telltale series used choices to engage players.
- *Exploration.* Opportunities to investigate a vast virtual space filled with mysteries and surprises encourage players to explore and discover different possibilities, consequently enhancing game engagement. Open world games like Fallout V or GTA often used this mechanism.
- *Different possible solutions.* With different possible solutions, players do not feel that they are forced by game designers to do any particular sequence of actions to solve a challenge but, depending on each player's creativity, they can come up with a unique solution. Excellent examples are chess and board games.
- *Not reward centered.* Designers should be cautious not to overuse rewards or other external motivators as the core motives for gaming engagement. The overuse of rewards and external motivators tends to diminish the intrinsic significance of the game essential elements and objectives and also to weaken the player's sense of autonomy because they are likely to feel controlled by these motivators and likely to withdraw at some point.

Competence

For competence, games should promote a sense of mastery. According to flow theory, the optimal experience is when the level of challenge and the user's skill level are both high. With such a balance, some users may enter their "comfort" zone in which they are likely to feel satisfied by their feeling of competency and self-esteem as witnessed through their accomplishments, while some users may enter a "stimulation" zone, in which they are likely to be stimulated to overcome even more difficult challenges. To design suitable challenges, Bushnell (http://nolanbushnell.com/) suggested designing challenges that are easy to learn but difficult to master.

It is crucial for designers to know the targeted user groups well. Such knowledge includes the backgrounds, understanding, familiarity, abilities, limitations and needs of various user types, in order to introduce appropriate challenges. But in HCI practice, it is often difficult to predict, determine, and understand users to this extent. Furthermore, in many cases, when users are given a very difficult task, they do not always come with the right set of skills. It is therefore important for HCI designers to understand the various mechanisms of game engagement that promote competence. The following game mechanisms help users achieve engagement from competence:

- *Ability to customize.* Each player has a different background, skill level and knowledge. Thus it is important to provide a flexible, user-adjustable interface. For example, games should offer a wide range of customization, such as the ability to adjust the level of difficulty, the ability to watch or skip tutorials, to customize preferences and features, and so on.
- *Feedback.* By learning from their actions and the consequences of those actions, feedback promotes player growth in skills and understanding, both of which are necessary for players to reach the optimal competence level.
- *Cheap failure.* The "retry pattern" allows players to retry when they fail the challenges. This allows gamers to learn quickly and become competent through failures, without unnecessary anxiety or stress, because these failures come at little or no cost. This mechanism turns player errors into positive learning experiences, allowing them to learn quickly and cheaply by encouraging them to review and practice the specific skill set required for each player to reach his or her optimal competence level.
- *Clear goal.* Allow users to concentrate fully on the task. By working towards a clear goal, the user is able to make clear priority judgments in the process and thus is less likely to be distracted. This focus is necessary for supporting the process of players' growth in their in-game ability.
- *Short and long-term goals.* Clear goals are important so that players can easily determine priorities and direction throughout the game process. The most important goal is usually the long-term objective of the game. Integrated short term goals act milestones as the game progresses providing and reinforcing the player's primary purpose and orientation and, when well designed, further motivating the player towards the goal.
- *Increasing difficulty.* A gradual increase in difficulty allows players to engage at an appropriate level and to practice and develop progressively the skills necessary to participate, compete, and complete each more difficult stage.
- *Tutorials.* Optional tutorials provide training so that players can become accustomed to the system.
- *Awards and trophies.* Awards and trophies deliver positive feedback that enhances the player's sense of competence and accomplishment. Effective reward systems follow a suitable reinforcement schedule (Skinner & Ferster, 1957). In any case, designers should be cautious not to overuse rewards or other external motivators as the core motives of game engagement as it can diminish the intrinsic significance of game objectives and also weaken the player's sense of autonomy.

Relatedness

For relatedness, players' drives should be acknowledged and socially connected. Thus, social mechanisms can be used to further enhance game engagement. Some social mechanisms include:

- *Collaboration*—collaborating in a team toward a shared goal can stimulate social engagement. This is often seen in highly popular multiplayer online battle arena game such as League of Legends or Counterstrike.
- *Leaderboard*—sharing and comparing one's achievements can promote indirect competition but designers should also be cautious not to alienate any groups of players due to their low skill levels or late entry.

- *Communication*—this gives players the ability to communicate and collaborate with other players, to share gaming experiences and achievements, or to discuss gaming strategies. This channel fosters each player's sense of belonging in a group or a gaming community.
- *Public space*—providing a public space where players can interact or gather together promote social interactions.
- *Representations*—providing players the ability to represent their own persona such as through avatars can foster players' identity.

Emotions

Emotional arousal describes a player's need to be occasionally aroused by various emotional stimuli. Players are likely to engage in a game that can elicit a *wide* range of *strong* emotional responses. Evocative highlights in a story trigger strong emotional involvement in players. In addition, the combination of a story with challenging decision making and multiple paths enables players to experience deeper levels of emotional engagement. Here are some game mechanisms that help promote emotional engagement:

- *Story.* Story stimulates curiosity and adventure instincts so that players want to continue engaging to find out what happens next.
- *Aesthetics.* Aesthetics (i.e., the look and feel of the game) influence user impressions and perceptions, thus users are likely to feel more positive and to be more engaged.
- *Realism.* Realism brings a game closer to reality; realism adds relationship and relevance between the gamer and the game itself, thus users tend to be more easily engaged and immersed. Nevertheless, designers need to be careful of the "uncanny valley" phenomenon that we mentioned earlier.
- *Role playing.* By assuming the role of a character in the game, by forming relationships within the game, and by going through the story as if they were living in that virtual world, the player forms a strong bond and emotional connection with the game. One good example is Pokémon Go, where people role play as a Pokémon master trying to catch all Pokémon.
- *Making difficult decisions.* Making difficult decisions that impact the overall story sequence and ending adds strong emotional commitment and involvement.
- *Character personalization.* In some games, gamers are allowed to personalize their characters' appearance. By personalizing their own characters, a sense of ownership and personal identification is encouraged; thus each gamer becomes more attached to his / her own character while playing.
- *Risks, uncertainty and competition.* Uncertainties in game outcomes are necessary to stimulate players to continue the gaming experience. Competition is a typical way to produce uncertainty in games.
- *Time pressure.* Giving players only a certain amount of time to achieve something can create a good amount of anticipation and suspense.

Closing Remarks

Our motivation in developing this survey was to provide a quick means of understanding the landscape of game engagement. Many nongame HCI researchers and practitioners are excited about emerging game-related areas, such as gamification and serious games, but there is a tendency to lack a clear, comprehensive understanding of game engagement. Providing a theoretical understanding can help practitioners make sense of game engagement, predict the likely outcome, guide the design process, and determine relevant variables to measure and monitor. It also broadens the possibility of discovering and exploiting new and existing techniques from games.

References

Ajzen, I. (1985). *Action control from cognition to behavior*. Berlin: Springer.

Andersen, E., Liu, Y.-E., Snider, R., Szeto, R., Cooper, S., & Popović, Z. (2011). On the harmfulness of secondary game objectives. In *Proceedings of the 6th International Conference on Foundations of Digital Games* (pp. 30–37). New York, NY: ACM.

Andersen, E., Liu, Y.-E., Snider, R., Szeto, R., & Popović, Z. (2011). Placing a value on aesthetics in online casual games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1275–1278). New York, NY: ACM.

Andersen, E., O'Rourke, E., Liu, Y.-E., Snider, R., Lowdermilk, J., Truong, D.,…Popovic, Z. (2012). The impact of tutorials on games of varying complexity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 59–68). New York, NY: ACM.

Baños, R. M., Botella, C., Alcañiz, M., Liaño, V., Guerrero, B., & Rey, B. (2004). Immersion and emotion: their impact on the sense of presence. *CyberPsychology and Behavior*, *7*(6), 734–741.

Bayliss, P. (2010). *Videogames, interfaces, and the body: The importance of embodied phenomena to the experience of videogame play* (Unpublished doctoral thesis). RMIT University, Melbourne.

Beenen, G., Ling, K., Wang, X., Chang, K., Frankowski, D., Resnick, P., & Kraut, R. E. (2004). Using social psychology to motivate contributions to online communities. In *Proceedings of the 2004 ACM Conference on Computer Supported Cooperative Work* (pp. 212–221). New York, NY: ACM.

Bellotti, F., Kapralos, B., Lee, K., Moreno-Ger, P., & Berta, R. (2013). Assessment in and of serious games: an overview. *Advances in Human-Computer Interaction*, *2014*(1), 1–11.

Benford, S., & Bowers, J. (1995). User embodiment in collaborative virtual environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 242–249). New York, NY: ACM.

Berkovsky, S., Coombe, M., Freyne, J., Bhandari, D., & Baghaei, N. (2010). Physical activity motivating games: virtual rewards for real activity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 243–252). New York, NY: ACM.

Bianchi-Berthouze, N. (2013). Understanding the role of body movement in player engagement. *Human–Computer Interaction*, *28*(1), 40–75.

Blythe, M., & Hassenzahl, M. (2004). The Semantics of Fun: Differentiating Enjoyable Experiences. In M. A. Blythe, K. Overbeeke, A. F. Monk, & P. C. Wright (Eds.), *Funology* (pp. 91–100). Norwell, MA: Kluwer Academic Publishers.

Boyle, E. A., Connolly, T. M., Hainey, T., & Boyle, J. M. (2012). Engagement in digital entertainment games: A systematic review. *Computers in Human Behavior*, *28*(3), 771–780.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Brown, E., & Cairns, P. (2004). A grounded investigation of game immersion. In *CHI'04 Extended Abstracts on Human Factors in Computing Systems* (pp. 1297–1300). New York, NY: ACM.

Caillois, R. (1961). *Man, play, and games*. Chicago, IL: University of Illinois Press.

Cheng, K., & Cairns, P. A. (2005). Behaviour, realism and immersion in games. In *CHI'05 Extended Abstracts on Human Factors in Computing Systems* (pp. 1272–1275). New York, NY: ACM.

Cialdini, R. (2008). *Influence: Science and practice.* Boston, MA: Allyn & Bacon.
Clark, D. (2007). *Games, motivation, and learning.* Sunderland: University of Sunderland / Caspian Learning.
Cox, A., Cairns, P., Shah, P., & Carroll, M. (2012). Not doing but thinking: the role of challenge in the gaming experience. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 79–88). New York, NY: ACM.
Csikszentmihalyi, M. (1990). *Flow : The psychology of optimal experience.* New York, NY: Harper Perennial.
Curran, N. (2013). *The psychology of immersion and development of a quantitative measure of immersive response in games.* Cork: University College Cork.
Davis, F. D., Bagozzi, R. P., & Warshaw, P. R. (1989). User acceptance of computer technology: A comparison of two theoretical models. *Management Science*, *35*(8), 982–1003.
Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological Inquiry*, *11*(4), 227–268.
De Kort, Y., IJsselsteijn, W., & Poels, K. (2007). Digital games as social presence technology: Development of the Social Presence in Gaming Questionnaire (SPGQ). In *Proceedings of the 10th Annual International Workshop on Presence* (pp. 195–203). Barcelona, Spain: International Society for Presence Research.
Desurvire, H., Caplan, M., & Toth, J. A. (2004). Using heuristics to evaluate the playability of games. In *CHI'04 Extended Abstracts on Human Factors in Computing Systems* (pp. 1509–1512). New York, NY: ACM.
Douglas, Y., & Hargadon, A. (2000). The pleasure principle: Immersion, engagement, flow. In *Proceedings of the 11th ACM Conference on Hypertext and Hypermedia* (pp. 153–160). New York, NY: ACM.
Ducheneaut, N., Wen, M.-H., Yee, N., & Wadley, G. (2009). Body and mind: A study of avatar personalization in three virtual worlds. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1151–1160). New York, NY: ACM.
Ermi, L., & Mäyrä, F. (2005). Fundamental components of the gameplay experience: Analysing immersion. *Worlds in Play: International Perspectives on Digital Games Research*, *37*(2), 37–53.
Federoff, M. (2002). *Heuristics and usability guidelines for the creation and evaluation of fun in video games* (Unpublished thesis). Bloomington, IA: Indiana University.
Festinger, L. (1954). A theory of social comparison processes. *Human Relations*, *7*(2), 117–140.
Freeman, D. (2003). *Creating emotion in games: The craft and art of emotioneering.* San Francisco, CA: New Riders.
Gee, J. (2008). Video games and embodiment. *Games and Culture*, *3*(3–4), 253–263.
Habgood, M. P. J. (2007). *The effective integration of digital games and learning content* (Unpublished thesis). University of Nottingham, Nottingham.
Hamari, J., Koivisto, J., & Sarsa, H. (2014). Does gamification work?—A literature review of empirical studies on gamification. In *47th Hawaii International Conference on System Sciences (HICSS)* (pp. 3025–3034). doi:10.1109/HICSS.2014.377
Hays, R. T., & Singer, M. J. (1988). *Simulation fidelity in training system design: Bridging the gap between reality and training.* New York, NY: Springer.
Huizinga, J. (1950). *Homo Ludens: A study of the play-element in culture.* Boston, MA: The Beacon Press.
Hunicke, R., Leblanc, M., & Zubek, R. (2004). MDA: A formal approach to game design and game research. In *Proceedings of the AAAI Workshop on Challenges in Game AI* (vol. 4, p. 1722). doi:10.1.1.79.4561
Isbister, K., Flanagan, M., & Hash, C. (2010). Designing games for learning: Insights from conversations with designers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2041–2044). New York, NY: ACM.

Ivory, J. D., & Kalyanaraman, S. (2007). The effects of technological advancement and violent content in video games on players' feelings of presence, involvement, physiological arousal, and aggression. *Journal of Communication*, *57*(3), 532–555.

Jalali, S., & Wohlin, C. (2012). Systematic literature studies: Database searches vs. backward snowballing. In *Proceedings of the ACM-IEEE International Symposium on Empirical Software Engineering and Measurement* (pp. 29–38). New York, NY: ACM.

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., & Walton, A. (2008). Measuring and defining the experience of immersion in games. *International Journal of Human-Computer Studies*, *66*(9), 641–661.

Kato, P. (2010). Video games in health care: Closing the gap. *Review of General Psychology*, *14*(2), 113–121.

Katz, E., & Blumler, J. (1974). *The uses of mass communications: Current perspectives on gratifications research.* Beverly Hills, CA: Sage.

Kim, K. J., Biocca, F., & Jeong, E. J. (2011). The effects of realistic controller and real-life exposure to gun on psychology of violent video game players. In *Proceedings of the Fifth International Conference on Ubiquitous Information Management and Communication* (pp. 49:1–49:5). New York, NY: ACM.

Klimmt, C. (2003). Dimensions and determinants of the enjoyment of playing digital games: A three-level model. In *Level up: Digital Games Research Conference* (pp. 246–257). Utrecht: Faculty of Arts, Utrecht University.

Klimmt, C., Rizzo, A., Vorderer, P., Koch, J., & Fischer, T. (2009). Experimental evidence for suspense as determinant of video game enjoyment. *CyberPsychology and Behavior*, *12*(1), 29–31.

LaViola Jr., J. J., & Litwiller, T. (2011). Evaluating the benefits of 3D stereo in modern video games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2345–2354). New York, NY: ACM.

Lazzaro, N. (2004). Why we play games: Four keys to more emotion without story. In *Game Developer's Conference* (pp. 1–8). Oakland, CA: XEODesign.

Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, *3*(2). doi:10.1111/j.1083-6101.1997.tb00072.x

Lucas, K., & Sherry, J. L. (2004). Sex differences in video game play: A communication-based explanation. *Communication Research*, *31*(5), 499–523.

Malone, T. W. (1980). What makes things fun to learn? Heuristics for designing instructional computer games. In *Proceedings of the Third ACM SIGSMALL Symposium and the First SIGPC Symposium on Small Systems* (pp. 162–169). New York, NY: ACM.

Mcmahan, A. (2003). Immersion, engagement, and presence: A method for analyzing 3D videogames. In M. J. P. Wolf & B. Perron (Eds.), *The video game theory reader* (pp. 67–86). New York, NY: Routledge.

Mekler, E. D., Bopp, J. A., Tuch, A. N., & Opwis, K. (2014). A systematic review of quantitative studies on the enjoyment of digital entertainment games. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems* (pp. 927–936). New York, NY: ACM.

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley. *IEEE Robotics and Automation Magazine*, *19*(2), 98–100.

Nabi, R., & Krcmar, M. (2004). Conceptualizing media enjoyment as attitude: Implications for mass media effects research. *Communication Theory*, *14*(4), 288–310.

Nacke, L. E., Grimshaw, M. N., & Lindley, C. A. (2010). More than a feeling: Measurement of sonic user experience and psychophysiology in a first-person shooter game. *Interacting with Computers*, *22*(5), 336–343.

Nakamura, J., & Csikszentmihalyi, M. (2002). The concept of flow. In C. R. Snyder & S. J. Lopez (Eds.), *Handbook of positive psychology* (pp. 89–105). New York, NY: Oxford University Press.

Norman, D. A. (2003). *Emotional design: why we love (or hate) everyday things.* New York, NY: Basic Books.

O'Brien, H. L., & Toms, E. G. (2008). What is user engagement? A conceptual framework for defining user engagement with technology. *Journal of the American Society for Information Science*, *59*(6), 938–955.

Oliver, M. B., & Raney, A. A. (2011). Entertainment as pleasurable and meaningful: Identifying hedonic and eudaimonic motivations for entertainment consumption. *Journal of Communication*, *61*(5), 984–1004.

Parker, J. R., & Heerema, J. (2007). Musical interaction in computer games. In *Proceedings of the 2007 Conference on Future Play* (pp. 217–220). New York, NY: ACM.

Phillips, C. A., Rolls, S., Rouse, A., & Griffiths, M. D. (1995). Home video game playing in schoolchildren: A study of incidence and patterns of play. *Journal of Adolescence*, *18*(6), 687–691.

Piselli, P., Claypool, M., & Doyle, J. (2009). Relating cognitive models of computer games to user evaluations of entertainment. In *Proceedings of the 4th International Conference on Foundations of Digital Games* (pp. 153–160). New York, NY: ACM.

Raney, A. A. (2003). Disposition-based theories of enjoyment. In J. Bryant, D. R. Roskos-Ewoldsen, & J. Cantor (Eds.), *Communication and emotion: Essays in honor of Dolf Zillmann* (pp. 61–84). Mahwah, NJ: Erlbaum.

Ravaja, N., Salminen, M., Holopainen, J., Saari, T., Laarni, J., & Järvinen, A. (2004). Emotional response patterns and sense of presence during video games: Potential criterion variables for game design. In *Proceedings of the Fourth International Conference on Foundations of Digital Games* (pp. 339–347). New York, NY: ACM.

Reinecke, L., Tamborini, R., Grizzard, M., Lewis, R., Eden, A., & David Bowman, N. (2012). Characterizing mood management as need satisfaction: The effects of intrinsic needs on selective exposure and mood repair. *Journal of Communication*, *62*(3), 437–453.

Reiss, S., & Wiltz, J. (2004). Why people watch reality TV. *Media Psychology*, *6*(4), 363–378.

Rossoff, S., Tzanetakis, G., & Gooch, B. (2010). Adapting personal music for synesthetic game play. In *Proceedings of the Fifth International Conference on the Foundations of Digital Games* (pp. 163–170). New York, NY: ACM.

Russell, J. (1980). A circumplex model of affect. *Journal of Personality and Social Psychology*, *39*(6), 1161–1178.

Ryan, R. M., Rigby, C. S., & Przybylski, A. (2006). The motivational pull of video games: A self-determination theory approach. *Motivation and Emotion*, *30*(4), 344–360.

Sanders, T., & Cairns, P. (2010). Time perception, immersion and music in videogames. In *Proceedings of the 24th BCS Interaction Specialist Group Conference* (pp. 160–167). Swinton: British Computer Society.

Schild, J., LaViola, J., & Masuch, M. (2012). Understanding user experience in stereoscopic 3D games. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 89–98). New York, NY: ACM.

Schutz, W. (1958). *FIRO: A three-dimensional theory of interpersonal behavior.* New York, NY: Holt, Rinehart & Winston.

Sherry, J., Lucas, K., Greenberg, B. S., & Lachlan, K. (2006). Video game uses and gratifications as predictors of use and game preference. *Playing Video Games: Motives, Responses, and Consequences*, *24*(1), 213–224.

Skinner, B. F., & Ferster, C. B. (1957). *Schedules of reinforcement.* New York, NY: Appleton–Century–Crofts.

Slater, M., Usoh, M., & Steed, A. (1994). Depth of presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, *3*(2), 130–140.

Sweetser, P. M., & Johnson, D. M. (2004). Player-centered game environments: Assessing player opinions, experiences and issues. In *Proceedings of Entertainment Computing* (pp. 321–332). doi:10.1007/978-3-540-28643-1_40

Sweetser, P., & Wyeth, P. (2005). GameFlow: a model for evaluating player enjoyment in games. *Computers in Entertainment*, *3*(3), 3–3.

Tajfel, H., & Turner, J. (1979). The social psychology of intergroup relations. In W. C. Austin & S. Worchel (Eds.), *An integrated theory of intergroup conflict* (pp. 33–47). Monterey, CA: Brooks / Cole.

Takatalo, J., Häkkinen, J., Komulainen, J., Särkelä, H., & Nyman, G. (2006). The impact of the display type and content to a game adaptation. In *Proceedings of the Eighth Conference on Human-Computer Interaction with Mobile Devices and Services* (pp. 17–20). New York, NY: ACM.

Tamborini, R., Bowman, N. D., Eden, A., Grizzard, M., & Organ, A. (2010). Defining media enjoyment as the satisfaction of intrinsic needs. *Journal of Communication*, *60*(4), 758–777.

Tocci, J. (2008). “You are dead. Continue?”: Conflicts and complements in game rules and fiction. *Eludamos. Journal for Computer Game Culture*, *2*(2), 187–201.

Vorderer, P., Klimmt, C., & Ritterfield, U. (2004). Enjoyment: At the heart of media entertainment. *Communication Theory*, *14*, 388–408.

Weibel, D., & Wissmath, B. (2011). Immersion in computer games: The role of spatial presence and flow. *International Journal of Computer Games Technology*, *2011*(6), 1–14.

Witmer, B. G., & Singer, M. J. (1998). Measuring presence in virtual environments : A presence questionnaire. *Presence: Teleoperators and Virtual Environments*, *7*(3), 225–240.

Yee, N. (2005). Motivations of play in online games. *CyberPsychology and Behavior*, *9*(6), 772–775.

Yun, C., Shastri, D., Pavlidis, I., & Deng, Z. (2009). O’ game, can you feel my frustration? Improving user’s gaming experience via stresscam. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2195–2204). New York, NY: ACM.

Zajonc, R. B. (1965). Social facilitation. *Science*, *149*(3681), 269–274.

Zillmann, D., & Bryant, J. (1985). Affect, mood, and emotion as determinants of selective exposure: In *Selective exposure to communication* (pp. 157–190). Hillsdale, NJ: Erlbaum.

Part VIII
Accessibility

30

Accessibility

Alistair Edwards

Introduction and Background

An important realization within the discipline of human-computer interaction is that the ability of people to use the technology varies. This realization has grown with the maturity of the discipline but it has also become more important as the spread of users has broadened. At one time computers were few in number and used only by specialists but now their use has become a daily necessity for most people in the developed world. (Indeed, as we will see later in this chapter, access to the Internet is increasingly being seen as a fundamental human right.) Small variations between users' abilities must be catered for—and much of the material in this book is concerned with matching technology to users—but there are also more fundamental differences in abilities that must be taken into account. Broadly, those are the differences that are so profound that they are generally classed as *disabilities*.

An important point to be highlighted in this chapter is how many of the technologies that are used in modern, ubiquitous technology were originally developed to meet the "special" needs of this group of users.

Firstly we need to set out some context in terms of accessibility needs. We present some particular approaches to accessibility and then look specifically at the question of access to the Web.

Disability and Interaction with Technology

Language

No one who is reading—or writing—a book can have any doubt about the fact that language is powerful. Around certain topics it has the power to inform but also to offend. Disability is a particular instance. Given the emotions that surround people's perceptions and self-perceptions, it is easy to offend. It is also easy to become vague and imprecise in an attempt not to offend. Language is also subject to fashion and evolution. A term that was seen yesterday as merely descriptive and informative can quickly become an insult.

The Wiley Handbook of Human Computer Interaction, Volume 2, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

Given the topic of this chapter, it is going to be necessary to refer to specific groups of people. I will attempt to do this in a way that is accurate and clear but also inoffensive. In so doing I have been guided by Sears and Hansen (2012).

Definitions

If we picture the classic image of a human using a computer, then it is probably a person sitting at a desk with a screen, a keyboard, and a mouse, so the ability to interact with technology can be affected by our abilities under simple, broad headings:

- sensory;
- physical; and
- cognitive.

Of course that picture is very much outdated, as we carry computers around in our pockets and bags and interactions with them take on new and varied forms. We all have different levels of abilities in these areas, but some people have particular difficulties, to the extent that they are identified as having a *disability*. A disability is defined in the UK Equality Act 2010 (the precise definition may vary in other jurisdictions, but they are all effectively similar) as “a physical or mental impairment that has a ‘substantial’ and ‘long-term’ negative effect on the ability to do normal daily activities.”

The definition is broad and open to a certain level of interpretation. Furthermore, while it is focused on the individual, it should be remembered that people do not operate in isolation; they interact with their environment. Thus a condition that may seriously affect a person’s ability to operate in a “hostile” environment but may have little effect in a well-designed, accommodating environment.

The perception that disabled people represent a small potential market has in the past meant that research on meeting their needs has often taken place more in the academic world than the commercial world. Academic researchers are not driven by the same profit motive and can relish the difficult challenges represented by accommodating extreme requirements. A major theme of this chapter is to highlight how many features in human-computer interaction were originally developed for users with disabilities—the perceived minority—that have now moved into the mainstream—to the benefit of all users. This has often happened because of the way the underlying technology has developed.

A turning point

In the early days of the development of user interaction, accommodating users with disabilities tended to be treated as an “optional extra.” That is to say, developers and manufacturers felt that catering for the majority users—those with no identifiable disabilities—was their priority, and that adaptations could be added for the others at some later date. So it was that users with disabilities were always playing catch up. Whenever a new system—or new version of a system—was released it was usually inaccessible to particular users. Then efforts would be put into finding ways of making the new device accessible and eventually released as some kind of add on. It often seemed that manufacturers were mostly content to leave the development of such adaptations to third parties—who had expertise in that area.

An important example is the *screen reader* for blind users. In the early days of the personal computer, interaction was mainly based on text. Text can be re-presented in the nonvisual forms of (synthetic) speech or braille. This led to the development of the screen reader, which was hardware—or latterly software—which could perform that translation, so making the computer accessible to blind users.

The problem became more difficult with the advent of the graphical user interface (GUI) whereby information presented on a screen was not purely textual, but nevertheless new screen readers were developed that made it possible for blind people to continue to use computers.

The Apple Macintosh was the first consumer personal computer with a GUI. Berkley Systems developed the *OutSpoken* screen reader (Edwards, 1995), but it was never very popular and Apple was not at all popular with blind users. Apple were seen by many as the villains in the technology. Screen readers were also developed by third parties for Microsoft Windows. Notably *Jaws* (Freedom Scientific, 2016) was the most popular, making Windows the system of choice for most blind computer users.

The point is that computer manufacturers seemed relatively oblivious to the needs of blind users and left it to others to develop the access technology. This meant that these developers—and hence their customers—were always behind. As new versions were released, blind users had to stick with the old technology until such time as their screen readers could be updated to work with the new versions.

The turning point was in 2005, when Apple released Mac OS X 10.4 (Tiger). This version of the operating system came with the *VoiceOver* screen reader *built in*. That is to say that finally blind customers were on an equal footing—at least in having access to new systems at the same time as everyone else. Later VoiceOver was also released on iOS (iPads and iPhones) and Apple had evolved from being the bête noir manufacturer to the brand of choice for many visually impaired users.

WebAIM regularly surveys screen reader users. A recent survey (WebAIM, 2015) shows that for screen-reader users, Windows is still the most used operating system (85.3%) so it is not surprising that Windows-based screen readers are also by far the most used (87.7%) with VoiceOver as low as 7.6%. However, with regard to mobile technology, 69.2% of respondents use a mobile phone, mobile handheld device or a tablet, and of them 69.6% use an Apple device with VoiceOver.

It is reasonable to suggest that the advent of VoiceOver does mark a watershed in terms of equality of access. Of course, that is *not* to say that it also marks the end to all the problems of disabled computer users. Notably much of the use of computers is to access the Web—and most of the Web remains to be poorly accessible, regardless of the technology used. (See below). Let us be clear, users with the profound disability of blindness are using the same technology as their peers, including the latest mobile technology.

Adapting to "disability"?

As mentioned earlier, the conventional human-computer interface, with a screen a keyboard and a mouse, was *constrained*—conveniently so for the human-computer interaction (HCI) developer. The scope of interaction within those constraints was narrow. Efforts could be expended on limited objectives such as redesigning the screen (with ever better, higher resolution displays), optimizing the keyboard—remember the important contribution of people like Card, Moran, and Newell (1983)—and utilizing the mouse (the GUI). Again, the work of people such as Stu Card and

Bill Buxton were important at the time (Buxton, 1986; Card, English, & Burr, 1978; Card, Mackinlay, & Robertson, 1990; Newell, 1995).

However, now the computer has escaped most of these constraints. The computer *is* your phone—and a mobile one. It is in your car. How, where and why we use computers has changed completely—as reflected elsewhere in this volume. The new devices open new and radical opportunities—but they also introduce new barriers and limitations. Operations that are easily achieved on a desktop machine can be quite hard to accomplish on (say) a small phone…in bright light, on the move (Figure 30.1).

This is an idea long promoted by Alan Newell (Newell, 1995). It is the concept that disability is not an isolated phenomenon: it is a product of people and their environment. We may all be disabled (or "handicapped," following the usage of World Health Organization, 2011) by our environment. Newell used the example of a soldier, "who can be blinded by smoke, deafened by gunfire, be mobility impaired by being up to the waist in mud, have poor tactile sensitivity and dexterity because of wearing a chemical warfare suit, and be cognitively impaired by being scared stiff" (Newell, 1995, p. 9).

That may seem somewhat extreme. More realistically, in the current context, a person with good manual dexterity (who can possibly type at 40 words per minute on a desktop computer) nevertheless may become a slow, "hunt-and-peck" typist on the

Figure 30.1 Ubiquitous computing. How, why and where we use computers has changed. Is this user "disabled"?

two-inch onscreen keyboard on their smartphone. They need help to overcome this limitation. Fortunately such help is at hand—thanks to the research that went on in the past to address the needs of users with identified disabilities.

We can briefly review the development of these techniques and technologies, up to the point that they have made it into the mainstream.

The ideal human-computer interface requires minimal effort—physical and cognitive—from the user. The goal of the discipline of HCI is to achieve that (this book is a handbook of human-computer interaction, so we will talk about that interface—but bearing in mind that although a lot of the technology we use currently does not resemble a "computer" as such, nevertheless many of the challenges to interaction are similar). This can be more difficult for some users for whom there is a greater mismatch between their abilities and the expectations built in to the interface. This has long been recognized and was crystallized by Don Norman in the form of the *Gulf of Execution* and the *Gulf of Evaluation* (Norman, 2002).

In some senses the whole of this book is about narrowing those gulfs. In this chapter we are concerned with doing this for particular users. For instance, a person with a motor impairment may not be able to *execute* input to a computer through a keyboard, and a blind person cannot *evaluate* output that is only available in a visual form. For the most part "assistive" technologies are required to bridge those gulfs.

Older people

Many disabling conditions are associated with increased age—which presents new challenges and opportunities for the technology. Most older people experience a decline in physical, sensory and / or cognitive abilities. Often no one of these degradations amounts to a "disability" in itself but in combination they do affect people's ability to undertake everyday activities, and (importantly) to live independently. Typically, an older person may need to wear eye glasses, will have a decline in short-term memory and may not be able to walk very far. In fact it is misleading, though, to say "typically" because there is no such thing as a typical older person. After much research, it can be concluded that the only thing that all older people have in common is that they have seen a lot of birthdays.

This makes it difficult to develop technologies that suit them. That which works well for the archetypal person just mentioned may be quite unsuitable for a contemporary who perhaps has better eyesight but a poorer memory.

There are often assumptions made about older people's reactions to technology. It may be assumed that they are suspicious of new technologies and find them hard to use. Again, this is a stereotype and may be true for some individuals but not for others. There are some reasons to believe that people's attitudes to technology are shaped by the technology that was prevalent in their formative years (i.e. their twenties, see Lim, 2010). For instance, those from a generation in which technology was something tangible such that when it broke down they could take it apart and repair it, may find software-based technologies that cannot be fixed with a screwdriver more difficult to use.

There have been suggestions that any discomfort with computer technologies will naturally be overcome as the population ages. People going into retirement now have used this technology in their work and so should be familiar with it. However, that is unlikely to be the case as the technology develops quicker than most people can keep up with. For instance, the person who has recently retired may be entirely comfortable with the keyboard, mouse and screen style of interaction—but not at all at ease with a touchscreen-based tablet. Again, this may be explained by their longer term experience of technology, rather than their most recent exposure.

Technologies

In this section we will review the basic technologies of interaction and see how they have evolved. The main theme is the way that technologies originally developed to meet the "special" needs of disabled users have merged into the mainstream, to the benefit of all users.

Input

Text Keyboards and their alternatives. As discussed above, inputting information into a computer implies bridging the Gulf of Execution. Much of that input is word based and the conventional approach is through the keyboard. Some people have difficulty using keyboards, and some cannot use them at all. Keyboards may be made (more) accessible through hardware adaptations including:

- *Keyguard.* This is usually a plastic sheet mounted on top of the keyboard with holes corresponding to the positions of the keys. The user presses through the hole. Having the guard in place means that there is less chance of accidentally hitting an adjacent key—even if the user has tremor in their hand. Furthermore, it is possible to rest the hand on the keyguard between key presses without pressing any keys (McCormack, 1990).
- *Pointing sticks.* Users who do not have sufficient manual control to type may type using a stick attached to another part of their body, such as the head or mouth. (Brodwin, Star, & Cardodo, 2004).

Software adaptations may also be used. Simple adjustments can be made such as changing the timing on keys. For instance, if users cannot remove their finger quickly enough they may get unwanted repeated copies of the letter. In such cases, the timing can be adjusted or the feature turned off altogether.

At another level, software can be used to replace the physical keyboard with an onscreen alternative. As long as the user can manipulate a pointing device (see below) they can use it to pick letters off the screen. The onscreen keyboard may be a simple analogue of a conventional qwerty keyboard, or may be a more innovative design aimed at just this kind of use, such as *Dasher* (Ward & MacKay, 2002).

An important point is that all of the above alternatives are much slower than a conventional keyboard. It is thus a good idea to use software to maximize the words input for the minimal number of keystrokes. This led to the development of *predictive input*. That is, based on the user's current input (and often on a history of their previous interactions) the system will predict the subsequent selections. Of course this kind of technology is familiar to most users of smart phones. Although it is the (deserved) butt of many jokes, it is generally a productive enhancement. After all, what is the keyboard on most smartphones but an onscreen keyboard that is hard to use because its keys are so small?

A paper investigating the theoretical limits of text prediction in English was published as far back as 1951 (Shannon, 1951). The earliest application of this to accelerating input for users with disabilities seems to be as far back as 1975 but now it is built into all phones.

Speech The keyboard may be seen as an entirely artificial form of input. A more natural, long-established form of communication is speech. To many, superficially, speech seems like the ultimate form of human-to-computer input. This is arguable. Speech is a very good channel of communication for a person talking to another person—someone with intelligence, affinity, and life experience, but these are qualities not found in most computers.

Once again, this technology had its birth in research into alternative forms of input and its application to some users with disabilities became apparent. Martin (1976) is one of the first references to this possibility. At that time there was a pessimistic tone regarding the practicality of speech recognition. Thus it was suggested that where for reasons of disability speech might be the only alternative, "limited vocabulary voice input systems" might be sufficiently good (Martin, 1976, p. 500).

The impairments that cause people not to be able to use keyboards often affect other forms of motor control—including that needed for speech. Speech input, therefore, tends to be used by people with conditions that only affect their manual control. Often these (ironically) are caused by overuse of keyboards. Variously referred to as *repetitive strain injuries* (RSI) or *work-related upper limb disorders* (WRULD) these can be caused by frequent and ergonomically poor use of the computer input devices, the keyboard, and mouse. An important aspect of treatment is usually to stop doing the thing that caused it in the first place—and speaking instead of typing is often part of that.

Early systems suffered from two usability problems. One was that in order to help the system to *segment* the input into separate recognizable words the speaker had to insert an audible—and unnatural—pause between each word. The second problem was that the systems had to be trained to each individual speaker before they could be used by that person. As anyone who has used speech-based systems such as Apple's *Siri* will attest, speech recognition has come on a long way in recent years. Siri (and its contemporaries, including Amazon's *Echo* and Google's *Assistant*). These offer speaker-independent, continuous speech recognition. They achieve this through advances in *deep learning*, based on *big data* (Deng, & Li, 2013; Yu, & Deng, 2014). However, in the context of this chapter, there are a number of points that need to be understood about these technologies.

Firstly, the main task of such systems is not necessarily to recognize every word of an utterance, but rather to extract meaning from it. This may seem like a harder problem. However, by simply picking out key words and using contextual information (the user's location, previous queries from that user and the like) it may be possible for the system to make a "good guess" at the intended meaning. This would contrast with someone using speech input for dictation—where picking out only a selection of "key" words will not do.

Another feature is that, although these apps are available on phones (with relatively low processing power) they do not run on the phone. Rather they are making use of the broadband capabilities and sending data to high-powered servers. It is on the server—with access to the large data mentioned above, and capable or running powerful algorithms—that the interpretation occurs.

The fundamental problem for dictation-style use remains to be the level of accuracy. Almost from the inception of the technology, manufacturers have been claiming accuracies of over 90%. For instance, Herman (1985) quoted 96% accuracy, while in 2016 Nuance claim "up to 99%" accuracy for their Dragon Dictate products (http://www.nuance.co.uk/index.htm). There are a number of points to note about such claims. The first is that they are rarely supported by credible evidence. Secondly, the definitions of "accuracy" are often unclear, and probably different in different instances. Thirdly, the difference from 100% is significant. "Up to" 99% is almost meaningless. What is its lower bound? Furthermore, even 1% errors can be frustrating.

Suppose a speech input technology can achieve 95% word recognition. That means that one in 20 words are misrecognized. In some applications that may be acceptable (again, digital assistants such as Siri might cope at that level) but for someone dictating text to their computer this can be frustrating. Misrecognition implies the need for editing to implement a correction. If users are controlling their entire interaction through speech, performing an edit can be laborious: steering the software's focus to the rogue word, selecting it, deleting it, getting it to recognize a word that it has just misrecognized, and then returning the focus to the original input point.

Again we have the situation that, for someone for whom speech is their only viable form of text input, less-than-perfect technology is better than none but life will be much better when 100% accuracy is achieved.

Pointing The second traditional form of input is the pointing device, conventionally the *mouse*. This is a fundamental part of the classical graphical user interface, affording the selection of objects on the screen. As discussed above, such mechanisms are sometimes further exploited to overcome some people's problems in using keyboards. All of this depends on the ability to point accurately with a device and to click buttons.

Alternatives to the mouse usually take the form of hardware. Trackballs and joysticks have the advantage for some people with motor control problems that the device remains stationary. This means that positioning movements can be separated from button presses.

Increasingly, mice are not used, but rather trackpads. These can be advantageous for some users, who find the direct, light touch easier, but they can cause problems for others. This can be exacerbated if complex, multifinger gestures (such as the *pinch*) are required.

Again, all these alternatives require manual control. Users who do not have that control may use other parts of their body. The head might be used, for instance (Guness, Deravi, Sirlantzis, Pepper, & Sakel, 2012). Commonly an arrangement of infrared transmitter, receiver and a reflector attached to the head can be used.

For someone who does not have the necessary head movement or control, eye gaze may be used. The development of eye-gaze technology is interesting. The idea of controlling technology by merely looking at it is attractive, but the reality is difficult. Like many of the technologies discussed in this chapter, it has often been considered only practical when the need for it is demanding—such as when there is little alternative. It is also very expensive. Mele and Federici (2012) provides a good review and Inclusive Technology (http://www.inclusive.co.uk/) lists currently available eye-gaze hardware.

All of these pointing mechanisms have the disadvantage that the pointing action and effect are separated: the user manipulates the mouse and the cursor moves, for instance. This was investigated in the early days of the graphical user interface, and the concept of *articulatory directness* was explored (Norman, 2002). This can be described as "how well the form and behavior of an input device (degrees of freedom, range of motion, discreteness of states) corresponds to the type of input values a user needs to express" (http://www.usabilityfirst.com/glossary/articulatory-directness/). Devices such as the mouse were seen as showing good articulatory directness and yet if one thinks about it even such a device involves an indirectness, a need to map from one plane to another. Lateral movements are direct: move the mouse to the left (or right) and the cursor moves to the left (right), but if one wants to move the cursor vertically *up* the screen, then one moves the mouse horizontally *away* from the body. Such movements are not so much direct as learned mappings.

However, the advent of the *touchscreen* has introduced a much more direct form of pointing. This is another example of a development that simplifies interaction for all users but is likely to have a more significant effect for people who find interaction more difficult—because of disability or age. This is surely one of the reasons that the modern innovation of the tablet computer seems to have been particularly successful for older users. For instance, in the United Kingdom, the number of people aged 65 and over who access the Internet via tablet computers increased from 17% in 2014 to 22% in 2015 (Office for National Statistics, 2014, 2015).

Whereas we have divided this section into a discussion of input and output, the touchscreen is probably the best example, which shows these are not separate: the movements of the finger on the screen would be meaningless without its visible reaction. So let us move on to look at the output component.

3.2 Output

Visual If we think again of the stereotypical human-computer interface, the remaining component is the screen. The basic technology has not changed much since the advent of bitmapped color displays in the 1980s. The details have changed in that screens have become bigger, flatter, thinner and—most importantly—higher resolution. This is not unimportant. Once again, these improvements may be welcome to the "average" user but decisive to some users who have visual impairments. The screen, as standard, is more visible but also will better support enhancements such as screen enlargement.

There are people for whom no screen is of any use—those who lack sufficient sight to use one. They must rely on nonvisual forms of information, which amounts to (synthetic) speech or braille.

Nonspeech sounds The "bandwidth problem" refers to the suggestion that it is possible to convey a lot of information in a visual form—more than it is possible to communicate in nonvisual alternatives. As discussed above, speech-based screen readers are successful alternatives to the visual screen—but they do not convey all the information on the screen at any time. There is scope, therefore, to add further nonvisual—and nonspeech—auditory information. There is no doubt that large amounts of information can be conveyed in (natural) nonspeech sounds, and yet there has been little advance in making more use of them in human-computer interfaces. There is more on this in Chapter 18, but in the context of this chapter it is probably sufficient to say that there is scope for much more use of nonspeech sounds for this group of users.

Haptics "Haptics" refers to sense and / or motor activity based in the skin, muscles, joints, and tendons. Within this chapter it is sufficient to note that the most developed form of haptic (cutaneous) communication is braille, and that this is an alternative form of (nonvisual) screen reader output.

The Web

Hitherto we have been looking at narrowing the Gulfs of Execution and Evaluation. According to Norman's model (Norman, 2002), these gulfs exist between the user and the technological artifact. There are instances in which the artifact itself may be designed in such a way that it facilitates—or hinders—the bridging of those gulfs. The Web is probably the best example of this.

The Web has rapidly become an essential part of everyday life in the developed world. We are quickly reaching a point where not to be able to access the Web is in itself a form of disability. Government, education, and commerce are moving to a point where in some cases the only way to interact is through the Web. As they reach that point, it is therefore imperative that their websites should be as accessible as possible. Often the main barrier to access may be simply economics: some people cannot afford the cost of the technology and the broadband connections. Their needs have to be addressed through social means—free access via libraries, for instance.

Once physical access has been achieved, then all of the barriers and solutions listed above come into play. Given the visual nature of the Web and most Web pages, one of the main barriers to access is for people with visual disabilities. The kinds of adaptations discussed above are equally applicable to Web access as to other use of the technology. However, there is another layer of adaptation represented by the design of the pages themselves.

Web pages can be designed in such a way that they are highly attractive visually but are hard or impossible to use via a screen reader. This has been known for a long time and much effort has gone into developing means of making pages accessible and encouraging designers to use these methods.

Much of the effort has been concentrated on *guidelines*, that is to say advice for page designers as to how to make their pages more accessible and to ensure that they are. Most prominent among these are the Web Accessibility Initiative (WAI) Web Content Accessibility Guidelines (W3C, 2012). These guidelines have now been through two iterations, the more recent version being dubbed WCAG 2.0. These embody the principles that websites must be *perceivable*, *operable*, *understandable*, and *robust*. For instance, under the *perceivable* heading, Guideline 1.1 is:

> Text Alternatives: Provide text alternatives for any nontext content so that it can be changed into other forms people need, such as large print, braille, speech, symbols or simpler language.

Compliance with some of the guidelines can be checked automatically, by software, such as *AChecker* (AChecker, 2011). For instance, as implied in Guideline 1.1, all images must have an alternative text ("alt") attribute. This is text that can be read out by a screen reader—which cannot display the image. However, not all the guidelines can be checked without human intervention. Take, for instance, Guideline 2.2:

> Enough Time: Provide users enough time to read and use content.

A software tool cannot generally measure the time that content will be displayed for, and, even if it could, what would be "enough" time? Compliance with this guideline is something that the developer ought to check manually. However, that opens the way to a continuing debate. There is a school of thought that guidelines are only a very small part of the solution, and that true accessibility can only be achieved by testing with real users (Kelly et al., 2007; Sloan et al., 2006). In this case the Web developer will probably be sighted and physically adept—and might have a very different perception as to how much is "enough" time to read and use content from (say) a screen reader user with poor manual dexterity.

The WCAG guidelines are assigned priorities and sites can be rated as to their level of achievement, the minimum being Level A conformance, up to Level AAA. In many countries failure to achieve at least Level A conformance is illegal—but there have been very few prosecutions. The best known example of a successful prosecution is probably the case brought against the Sydney Organising Committee for the Olympic Games, whose site was found to be inaccessible (Australian Human Rights Commission, 2000). It was found that the Committee had engaged in unlawful discrimination. It was ordered to render its website accessible including all text on all images and image map links on its website; to provide access to the Index of Sports from the schedule page; and provide access to the results tables. Failure to comply would have left the option to press for compensation.

This case was brought under the Australian Disability Discrimination Act 1992. Similar legislation exists in other countries, notably the Americans With Disabilities Act (ADA) in the United States (Bick, 1999). There are moves to extend legislation to the concept of a Bill of Internet Rights, promoted particularly by Tim Berners-Lee. It is suggested that access to the Internet has reached the stage of necessity, of being a human right. If this is to be adopted, then clearly all barriers to access will have to be addressed, including those associated with disabilities.

Web accessibility has hitherto been given a low priority by most Web developers (Disability Rights Commission, 2004) but there is an increasing awareness of the commercial and legal pressure to achieve accessibility—although there is still a long way to go. It seems likely, though, that if accessibility is raised to the level of a human right, then the pressure to achieve it will be increased.

The Future

In this chapter we have traced the status of research and development on access in challenging circumstances from a minority, specialist interest to mainstream, everyday application. What has changed is not the user but the technology and how we use it.

The chapter commenced with an almost apologetic discussion of "language," in awareness of the sensitivity of many people with regard to descriptions of conditions considered to be "disabilities." It is perhaps idealistic to envisage a time when such discussions will be redundant, when the use of such terms will be almost meaningless. It will take a lot more development, not to say a revolution, for this to become the case generally in society, to the point that disability could become an almost meaningless concept with regard to access, education, employment, and so forth. However, it may well be that there will be a general realization that people's abilities lie on a spectrum and that elements of their environment (such as computers) can be designed to accommodate people almost anywhere on that spectrum. Yes, this is probably idealistic—and yet this chapter has shown that advances have been made in that direction.

For instance, it is now possible for a blind person to buy precisely the same—and most up-to-date—technology as their sighted colleague and use it immediately. At the same time, the sighted person will also find that technology easier to use, because of features originally developed and designed to assist users identified as disabled. There is no reason not to hope and expect this trend to continue. For instance, at the moment "multimodal" interaction with computers is seen as something specialized, yet human-human interaction is possible between people of all kinds of abilities largely because it is inherently multimodal (with built-in redundancy). As multimodality becomes the norm in human-computer interaction, we can expect that it will become increasingly easy to use the technology—for all users.

References

AChecker. (2011). Web accessibility checker [Online accessibility checker]. Retrieved from http://achecker.ca/checker/index.php

Australian Human Rights Commission. (2000). *Bruce Lindsay Maguire v Sydney Organising Committee for the Olympic Games.* Retrieved from https://www.humanrights.gov.au/maguire-v-socog

Bick, J. (1999). Americans with Disabilities Act and the Internet. *Albany Law Journal of Science and Technology*, *10*(2), 205–227.

Brodwin, M. G., Star, T., & Cardoso, E. (2004). Computer assistive technology for people who have disabilities: Computer adaptations and modifications. *Journal of Rehabilitation*, *70*(3), 28.

Buxton, W. (1986). There's more to interaction than meets the eye: Some issues in manual input. In D. A. Norman & S. W. Draper (Eds.), *User centred system design: New perspectives on human-computer interaction* (pp. 319–337). Hillsdale, NJ: Lawrence Erlbaum Associates.

Card, S., English, W., & Burr, B. (1978). Evaluation of mouse, rate-controlled isometric joystick, step keys and text keys for text selection on a CRT. *Ergonomics*, *21*(8), 601–613.

Card, S., Mackinlay, J. D., & Robertson, G. G. (1990). The design space of input devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 117–124). New York, NY: ACM.

Card, S., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction.* Hillsdale, NJ: Lawrence Erlbaum.

Deng, L., & Li, X. (2013). Machine learning paradigms for speech recognition: An overview. *IEEE Transactions on Audio, Speech, and Language Processing*, *21*(5), 1060–1089.

Disability Rights Commission. (2004). *The Web: Access and inclusion for disabled people.* Retrieved from https://www.city.ac.uk/__data/assets/pdf_file/0004/72670/DRC_Report.pdf

Edwards, A. D. N. (1995). Outspoken software for blind users. In A. D. N. Edwards (Ed.), *Extra-ordinary human-computer interaction* (pp. 59–82). New York, NY: Cambridge University Press.

Freedom Scientific (2016) The world's most popular Windows screen reader [Software]. Retrieved from http://www.freedomscientific.com/Products/Blindness/JAWS

Guness, S. P., Deravi, F., Sirlantzis, K., Pepper, M. G., & Sakel, M. (2012). Evaluation of vision-based head-trackers for assistive devices. In *Engineering in Medicine and Biology Society (EMBC), Annual International Conference of the IEEE* (pp. 4804–4807). Piscataway, NJ: IEEE.

Herman, P. A. (1985). The Effect of Repeated Readings on Reading Rate, Speech Pauses, and Word Recognition Accuracy. *Reading Research Quarterly*, *20*(5), 553–565.

Kelly, B., Sloan, D., Brown, S., Seale, J., Petrie, H., Lauke, P., & Ball, S. (2007). Accessibility 2.0: people, policies and processes. In *Proceedings of the 2007 International Cross-Disciplinary Conference on Web Accessibility (W4A)* (pp. 138–147). New York: ACM.

Lim, C. S. C. (2010). Designing inclusive ICT products for older users: Taking into account the technology generation effect. *Journal of Engineering Design*, *21*(2–3), 189–206.

Martin, T. B. (1976). Practical applications of voice input to machines. *Proceedings of the IEEE*, *64*(4), 487–501. doi:10.1109/PROC.1976.10157

McCormack, D. J. (1990). The effects of keyguard use and pelvic positioning on typing speed and accuracy in a boy with cerebral palsy. *American Journal of Occupational Therapy*, *44*(4), 312–315.

Mele, M. L., & Federici, S. (2012). A psychotechnological review on eye-tracking systems: towards user experience. *Disability and Rehabilitation: Assistive Technology*, *7*(4), 261–281.

Newell, A. F. (1995). Extra-ordinary human-computer interaction. In A. D. N. Edwards (Ed.), *Extra-ordinary human-computer interaction: Interfaces for users with disabilities* (pp. 3–18). New York, NY: Cambridge University Press.

Norman, D. (2002). *The design of everyday things.* New York, NY: Basic Books.

Office for National Statistics. (2014). *Internet access—households and individuals: 2014.* Retrieved from http://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2014-08-07

Office for National Statistics. (2015). Internet access—households and individuals: 2015. Retrieved July 27, 2016 from http://www.ons.gov.uk/peoplepopulationandcommunity/householdcharacteristics/homeinternetandsocialmediausage/bulletins/internetaccesshouseholdsandindividuals/2015-08-06

Sears, A., & Hansen, V. L. (2012). Representing users in accessibility research. *ACM Transactions on Accessible Computing (TACCESS)*, *4*(2). doi:10.1145/2141943.2141945

Shannon, C. E. (1951). Prediction and entropy of printed English. *Bell System Technical Journal*, *30*(1), 50–64.

Sloan, D., Heath, A., Hamilton, F., Kelly, B., Petrie, H., & Phipps, L. (2006). Contextual web accessibility-maximizing the benefit of accessibility guidelines. In *Proceedings of the 2006 international cross-disciplinary workshop on Web accessibility (W4A): Building the mobile web: rediscovering accessibility?* (pp. 121–131). New York, NY: ACM.

W3C. (2012). *Web content accessibility guidelines (WCAG) overview.* Retrieved from http://www.w3.org/WAI/intro/wcag.php

Ward, D. J., & MacKay, D. J. C. (2002). Fast hands-free writing by gaze direction. *Nature*, *418*(6900), 838.

WebAIM. (2015). *Screen Reader User Survey #6 results.* Retrieved from http://webaim.org/projects/screenreadersurvey6/

World Health Organization. (2011). *World report on disability.* Geneva: World Health Organization.

Yu, D., & Deng, L. (2014). *Automatic speech recognition: A deep learning approach.* Berlin: Springer.

31

Designing for and with People Living with Challenging Circumstances

Kellie Morrissey and John McCarthy

Starting Position

Recent directions in HCI research have intersected with psychological and medical research in order to create technologies for people living with illnesses, disabilities and disorders, as well as people who fall outside the "normal" spectrum of use—for instance, elderly people or young people with physical disabilities. These directions in research have been broadly described as designing for vulnerability or designing for accessibility or inclusion (Carrington, Hurst & Kane, 2014; Vines, McNaney, Lindsay, Wallace, & McCarthy, 2014), and the general approach taken in this research is to use the technology that is designed in order to circumvent the issue that is put in place by the illness or disability itself. Many of these technologies are useful, important, or even emancipatory for the people whom they serve—for instance, prosthetic and robotic limbs for amputees and motorized wheelchairs offer a freedom of movement and a quality of life that had been denied to their users before.

This becomes more complex when we try to design technologies that seek to "make up" for what are perceived as deficits or dysfunction in mental abilities. For example, wearable alarms that signal caregivers when an elderly person is likely to fall or to leave the house may offer important protections for the safety of the elderly person; however, is it ethical to treat the actions of a person as needing to be monitored and alarmed? What must it feel like to wear a GPS, which makes you beholden to your family well into your adult life? Similarly, many attempts at creating video games for children with learning disabilities prioritize instrumental actions and end up ignoring the need for such games to be aesthetically and sensually pleasing to their players as well as challenging. In short, designing for vulnerability and disability, though its intentions may be good, run the risk of dehumanizing participants or—ironically—reducing them to their disability or to the set of circumstances that they find themselves facing.

A key move that we suggest in this area is the move from *disability* to *different abilities*. Many people who face disabilities or challenging circumstances find ways of adapting to these circumstances in order to live their lives in the ways that they need (Freedman et al., 2015). The common example that is given here is the propensity of

The Wiley Handbook of Human Computer Interaction, Volume 2, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

people who are born or become blind to experience a honing of existing senses—for example, a sharpened sense of hearing or smell (Catteneo & Vecchi, 2011). In these ways, their abilities can surpass those of the "normal population." Some different abilities are somewhat learned *in situ*—again, for instance, people who are born or become deaf often learn to read lips or to communicate using sign language. The abilities that we prioritize as a society are those that we have made necessary in order to navigate the society that we have built—in this way, ability is something socially constructed (Anastasiou & Kauffman, 2011). A move to considering disability as different abilities is one that opens up our design sensibilities to the spectrum of human potential inherent in every user. Here, we differentiate our stance from that of "inclusive" design—our starting point is instead oriented to potentials and possibilities in a spectrum of experience than an attempt to configure existing design for people whose abilities are different from ours.

But what about people who are not living with disabilities, but instead face challenging social and psychological circumstances in which digital design could play a part or even help to alleviate? Or what about the intersection of the two—those living with illness or disability and who find their circumstances worsened or complicated by living in a society that positions them on the outside? A subset of design research studies in the recent past have addressed such populations, broadly calling the studies "design for and with *vulnerable* people"—perhaps most notably a workshop for CHI 2013 (Vines et al., 2014), in which the authors of this chapter were engaged as participants and organizers.

In the years following, there has been a heightened awareness of what it might mean to position our participants as vulnerable, both in a larger context in the literature and in our own professional practice. The term has several concurrent issues. On the surface, the most immediate issue is that of its broadness. Schroeder and Gefenas (2009) acknowledge the problem of a lack of a definition of the term "vulnerability," suggesting instead that it is something that we can more easily see in action than define in academic terms. The authors then proceed from both a commonsense and research ethics view to arrive at the following definition:

> To be vulnerable means to face a significant probability of incurring an identifiable harm while substantially lacking ability and/or means to protect oneself (p. 117).

However, they then go on to discuss the broadness with which the term "vulnerability" has been applied in research ethics of late; a point also made by Woelfer (2014), who questions whether we should consider homeless people (particularly homeless youths, the population she carries out research with) "vulnerable" or, instead, "susceptible." Woelfer argues that the term "vulnerable" has been too broadly applied in research so that it is effectively meaningless. For instance, most research guidelines would term young people as vulnerable, regardless of their housing situation. Homeless youths, therefore, are susceptible, meaning that they have already suffered harm and may be susceptible to suffering further harm by virtue of their situations. Woelfer then describes her HCI research with this population under the headings of *long-term commitment* (little is known about design with this population, necessitating such a commitment), *participation* (to entangle both the technology, the design processes, the population and the situations together), *precaution* (technological systems almost always mean an intervention in a participant's life, which may do either harm or good), and *method* (diversification of methods in order to capture different modalities of experience). Although Woelfer's use of susceptibility as a stand-in for vulnerability, we argue, is still too vague, her themes of commitment, participation, precaution and method will echo throughout this chapter.

Though the breadth with which the term is applied is sometimes problematic, our own discomfort with the word "vulnerable" as we use it to apply to our participants is in its politics. To term a whole population vulnerable, although it is often an attempt to acknowledge the complexity of experience, often does the opposite. It imbues a view of the population as sufferers, as victims, and indeed makes that the primary characteristic of the population (Clarke & Chenoweth, 2006). Although this is often done in order to protect certain populations above others in research it can arguably do a disservice to the people whom we work with. Moreover, it constrains the subject matters we explore—researchers are less likely to explore issues of creativity, fun, play, expression, performance, and personhood with a population who has been demarcated as "vulnerable." Instead, the term risks elevating the researcher/designer as a "savior," or someone who can fix a supposed problem in the situation, while the population we work with becomes little more than the object of our research.

Returning to Woelfer's point on the back of our discomfort with the term "vulnerable"—if everyone is vulnerable in many different ways, perhaps it makes more sense to ask what it means to not be vulnerable. A term that is often contrasted with "vulnerability" in applied psychological research is "resilience." Resilience is defined by Ungar (2008), codirector of the Resilience Research Centre, as the following:

> In the context of exposure to significant adversity, resilience is both the capacity of individuals to navigate their way to the psychological, social, cultural, and physical resources that sustain their wellbeing, and their capacity individually and collectively to negotiate for these resources to be provided in culturally meaningful ways (p.225).

Resilience is not the opposite of vulnerability, but rather a response to vulnerability. When facing adversity, a vulnerable person is resilient if they have the capacity to navigate to a situation where their wellbeing is sustained. Moreover, Ungar's definition takes into account the possibility of vulnerable groups to act collectively and to negotiate for resources to be provided in ways that are valuable and make sense to them.

In this way, we can see that people's response to challenging situations are very much like the predisposition of certain senses to become elevated in blind or deaf people—and though this effect is due in part to the neuroplasticity of the brain (Bavelier, Dye & Hauser, 2006), it too is a learned behavior or an adaptation to circumstances. These adaptations may take many forms. Whereas some facing difficult circumstances may turn to family, friends, education in order to improve or alleviate their situation, some may carry out only minute actions that nevertheless represent a form of sense making or resilience in their lived experience. This chapter will go on to detail research work with people with dementia, whose forays into creative activity, we argue, constitutes a form of resilience in their experience.

Ungar's definition resonates with us. It is within this definition that we position ourselves at the outset of this chapter, namely, that HCI systems designed for and with people facing challenging situations should either (a) enhance these individuals' capabilities to maintain their wellbeing or (b) be a part of the (collective) negotiation process for these resources to be provided. Moreover, we propose that the technologies that we create in HCI should not be created solely for people living with challenging situations. This risks a double exclusion. Rather HCI design takes into account the diversity human action, experience and imagination such that designing for the creativity of people with dementia helps us understand and design for all creativity.

A focus on experience is key to appreciating the value of difference and diversity, and technologies that promote resilience also enrich our experiences in many different ways.

Experience and Variation

Every experience any of us has in our lives, whether we consider them it as a tacit, fleeting moment (the feel of grass against bare legs on a sunny day) or a larger chunk of time (remembering your friend's wedding day), is necessarily experienced from our own particular points of view. The point of experience and perspective is one that will return strongly throughout this chapter.

Our orientation toward experience has been informed by the work on aesthetic experience by Dewey (1934), who held that experiences and interactions are not just within or outside of the person but instead happen between the person and the world. Much of the research we encounter in this chapter focuses not necessarily on places or problems as sites for design but instead looks at how relationships (both intimate and broader community and socio-political relations) can function as sites for design. We are particularly interested in the view that interactions between two (or more) people are not back and forth, instrumental procedures but instead constitute a shared space between the participants in the conversation.

McCarthy & Wright (2004) describe a way of thinking about technology and a way of designing (Wright & McCarthy, 2010) that pays attention to experience above all else. When they (and we) talk about experience, they are talking about lived experience, and prereflective, tacit experience. They also talk about the ways in which people understand this experience, make sense of it, and explain it to others. Their approach to conceptualizing technology puts an emphasis on the emotional and sensory qualities of our immediate experiences.

Technology is so often conceptualized as something that is solid and apart from human experience—however, as McCarthy and Wright detail, it is now a part of our everyday lives, and, if sensitively designed, has the potential to play a part in the resilience-making process of vulnerable populations to become a resource that is used to maintain both stasis and wellbeing in turbulent or challenging situations. When designing for and with these populations, it can be difficult to access their experience in a way that allows us to use these contributions in effective ways in our designs. For example, in working with children with learning difficulties or with migrant women who have faced domestic violence we may face a host of communication difficulties that renders the normal user-centered process of design almost unworkable. Even sensitive, qualitative methods like interviews can be troublesome in contexts such as these.

To a degree, this is a challenge that proceeds from experience-centered design (Wright & McCarthy, 2010)—to what extent can the contributions of our participants be said to be equal to those of the research team in our shared design process? This is made more important when the population we are working with is a vulnerable population whose experiences and ways of communicating may differ from our own. Ensuring the perseverance of the voice and the needs of our participants in any resultant design work is imperative in order to ensure that the work we carry out is true to their own lived experience, and thus will be something that can be used, and indeed enrich, their lives.

Experience can also be a critical point in design research. We turn to work by McCarthy & Wright (2006), which positions the turn to experience as necessarily critical due to the ambiguities and attention to lived experience and felt life within these projects: within this uncertainty, experience "finds its critical edge." Immersed in our everyday, prosaic experiences as we build relationships, engage ourselves in design, and make an impact upon the world around us, reified and unworkable concepts fall away in the face of what we know, in our own experience, to be true for us. Thus, paying attention to the experience of both the researcher and the researched in design research is likely to yield both ethical and critical results.

Applied Areas

The following sections will describe briefly a number of different design projects with different populations, all living with illness, disability or otherwise facing difficult times in their lives. First we will discuss the challenges of living as a homeless person, and take a look at the appropriation of technologies that this population uses to maintain its own wellbeing, as well as a series of studies that detail the complexities inherent in carrying out a research process with a group of people whose circumstances change daily.

We will then discuss projects with people with dementia, focusing on two main approaches to designing with this population—one that pays particular attention to the deficits inherent to the disease, and one that prioritizes the abilities that are still present in participants.

Finally, we will briefly visit a participatory project that investigated expressive approaches to dealing with the aftermath of domestic violence for a set of migrant women in the north of the United Kingdom.

Following these area summaries, we will look at three overarching themes that run through all of these projects. We will close this chapter with some critical points on designing with people facing challenging circumstances as well as solidifying some of the more salient themes running through these projects as ways forward when designing with this population.

Homelessness

The causes of homelessness are manifold and can be either structural (poverty, unemployment, eviction, crime, antisocial behavior, and debt), or individual (addiction, family breakdown, domestic violence, leaving a state institution, or mental health issues) (Anderson & Christian, 2003). Whatever the cause, homelessness can have deleterious effects on the individual's health (cold injuries, physical and sexual assault, nutritional deficiencies) and wellbeing (loss of self-esteem, loss of will to take care of oneself, increased possibility of drug dependence). While the problems that cause homelessness are often social in nature and the solutions are undoubtedly rooted in social change on a large scale, the design and use of technological systems can improve the situations of many homeless people through the design of technologies that allow easy and effective communication and which facilitate learning about issues of shelter and health.

Le Dantec and Edwards (2008) carried out interviews with 13 homeless people on their interactions and experiences with technology. Several themes emerged, the primary one focusing on the importance of staying connected to friends and family—often over a distance, as some participants were displaced from their original homes by Hurricane Katrina:

> I haven't seen my momma since Katrina [2005]…When I was in Houston, the Red Cross…had a system to put our year, our date and the last address…Where ever my mother was at, that's how they tracked her down. Though when I called her she was on the voice thing…I never talked to her I just heard her voice on the thing.

The stress of becoming newly homeless added to a sense of estrangement from their usual social networking was remarked upon by several participants. In the absence of their "old" social network, new connections were sought out and forged, usually via their caseworker, a central figure in their social network. The caseworker is an integral facet in the lives of these people, facilitating access to medical services and acting as a mediator in the cases of participants who were illiterate or who suffered from mental illnesses.

Health and medication were issues identified by Le Dantec and Edwards as possible sites for technological intervention—with homelessness increasing the chances that persons will suffer from physical and mental illnesses, homeless persons often face difficulties in managing their health and medication. One participant, whose formal education finished around age 9, was unable to read as a result, and in this way faced a particularly difficult problem:

> Well, you see by not knowing how to read I go uh, what I do, I know the pills…[and] I got a little sack, a little medicine sack. I have ten bottles of pills so I dump em all out on the bed and…every time I take one out the bottle, I put the bottle in the sack so I can't go wrong.

While these devised strategies clearly demonstrate practicality and ingenuity, this practice is at best inefficient and at worst can be very damaging. Technological interventions that relied on verbal and aural interfaces instead of reading could alleviate the stress and danger of situations like this.

Tied up with the notions of self-care and alienation from social networks is an issue that we visited in the previous section—the issue of self and identity. Becoming homeless may pose a challenge to one's own self-concept; this is a space in which our sense of self may change completely or indeed, may be lost—although it is important to acknowledge that, for some people, it may be a life choice that is the lesser of many other evils.

> It's one thing being homeless but it's another thing…disappear[ing] from the face of the earth. And that's the biggest danger for homeless people. That's the hardest thing to manage, is when you get disconnected.

Homeless persons in this study used technological options to navigate and present their identities. While some felt that their ownership and use of cell phones connected them to their "old" circles ("they know if I got my cell phone I must be doing alright"), others still used free Internet services provided by the library to create and maintain MySpace accounts. For these homeless persons, the use of such social networking sites was mostly communicative and was used to contact and keep in touch with friends and family.

One project that aimed to address these needs—given the lack of availability of such technology for many persons without a home—was the #Patchworks Catalyst subproject (Southern et al., 2013), which investigated, via a series of codesign workshops, homeless participants' interactions with technology. These workshops led to a series of observations by the researchers about the nature of running such workshops, including using personas to deflect any questions that are too personal or too painful, and in particular the conflicting schedules many participants had (attempting to secure accommodation in different places daily, appointments with care representatives, collecting welfare), which caused them to drop out of workshops one week but attend and participate enthusiastically the next. The resultant prototype—#Pat—was a reminder system for upcoming appointments that use RFID technology to identify the user. #Pat was a low-fi but distributed technology, available at several points around town and presented the user with a receipt slip detailing any upcoming appointments to avoid the frequent use of caseworker time to be reminded of upcoming appointments. Although #Pat proved useful for the participants involved, we would like to highlight the lessons learned when trying to run workshops with the homeless participants, and will return to this in a later section.

The use of technologies and services by homeless participants in the above studies may seem scattered and disorganized, it can also be seen as the creative use of a series of systems and ICT products that help to organize their lives. Later in this chapter, we will discuss this creative use of available ICT, even if it is not state of the art ICT, as a feature of these participants' resilience.

Dementia

In this section, we will discuss some design research that has been carried out for and with people living with dementia and their carers. Dementia is an umbrella term for a variety of symptoms, but sufferers most often suffer from problems with short-term memory, attention, language and planning, and problem solving. Dementia is a health priority in most First World nations, and one way to alleviate the burden of dementia, both from a cost perspective and from a care perspective, is through the design of effective technologies. The term often used for such technologies is "assistive technologies," which refers to "any device or system that allows an individual to perform a task that they would otherwise be unable to do, or increases the ease and safety with which the task can be performed" (Royal Commission on Long Term Care, 1999).

There are several different types of assistive technologies. Many devices focus on the deficits in short-term memory, which are most characteristic of the condition—two examples of such memory aids are Memory Glasses and Microsoft's SenseCam (Bharucha et al., 2009), which aim to ameliorate retrospective forgetting and prospective forgetting by the user in turn. Physical and environmental sensors have also been deployed in order to remotely monitor the movements of people with dementia, and fall and agitation detection technologies for people with dementia show some promise (Williams, Victor, & McCrindle, 2013), while ambient living technologies have also been deployed in order to afford the person with dementia living at home some independence and safety in everyday tasks.

While the above discussed technologies have clear benefits for the physical safety of people with dementia, there are a subset of needs that are not catered for by the design and use of these technologies. These needs are somewhat harder to explicate, particularly for people with dementia, but they include the retention of a sense of self, the felt experience of day-to-day life as they remember it, reminiscing about happy times that may be far in the past, and a sense of independence and ability that goes beyond the ability to make a cup of tea unaided or recall the events of the past day (Ryan, Bannister, & Anas, 2009).

How do you design for and support these difficult-to-grasp needs? A first step towards doing so is found in the attempt to understand their experience in the ways in which we described earlier. However, in dementia—a condition that often sees communication styles change and verbal communication wane—accessing the experiences of people with dementia may need to be carried out in an oblique way. In order to complete our review of technologies for ageing persons and persons with dementia, we will look at two sets of studies—one by Kellie Morrissey, the lead author of this chapter, who looked at designing for musical interactions with people with dementia living in care, and the other by Jayne Wallace, who approached these difficult design questions in a case study focusing on an aging couple, wherein the wife had recently received a diagnosis of dementia.

People with dementia living in care constitute an "in-between" community—one that does not necessarily belong "on the inside," where they live, but through circumstance and the progression of their condition, cannot belong on the outside either (Probyn, 1996). Moreover, although this population lives with others, often for many years, in very communal settings, they can also suffer from a lack of socialization (which can hasten the progression of many dementias) as well as social isolation. Initially interested in the ways in which people with dementia take part in creative activities as a way of understanding possible future design sessions with the population, Morrissey (2015) and Morrissey and McCarthy (2015) carried out a longitudinal ethnography into life in care homes for people with dementia. Initially investigating by way of creative workshops (i.e., baking, art), she instead became interested in the ways in which residents of the care home would take part in music "sessions" or moments of music that, unlike the baking and art sessions, they instigated themselves and which were extremely active.

During these music sessions, one prop in particular was useful—that of a large, shimmering piece of cloth that residents, sat in a line, would hold together and use to shimmy and sway in the same directions while listening and singing along to music. However, this swathe of cloth had some smaller issues—it tended to "trap" residents who didn't wish to take part in the middle of the swaying cloth, which had potentially serious ramifications for those who might want to arise and go to the bathroom. Moreover, although it was generally seen as good fun for participants, it prevented some residents from improvising or carrying out their own creative moments within the flow of the session as they usually had no choice but to use the swaying piece of cloth.

We wondered if there was a similarly simple design that we could create and prototype together that would preserve a sense of community or "doing together" but also enhance the individual creativity of residents. With this in mind, we had several design features that we wished to preserve:

- providing interactions for residents;
- providing familiar but engaging interactions;
- easy to understand and operate for residents;
- allowing space for community building;
- rhythmic / melodic (or otherwise enhancing existing music sessions);
- would not "trap" residents into a particular activity, whether by requiring a long amount of time or engendering physical barriers;
- the potential for carers to involve themselves with sessions;
- the potential for future iterative codesign with residents;
- easy to operate within the confines of a care center.

Our resulting design was titled SwaytheBand—a simple set of Bluetooth-linked (to a central computer) PS Move controllers. During a song, the participants hold the controller as the LED light at its top changes color in time to the beat of the music. The controllers are connected to a central computer via Bluetooth (using the PS Move API), which dictates the color and timing. This configuration is able to support up to eight participants with the computer being additionally responsible for playing the music.

Early field trials with the design have been promising. Those using the batons use them to sway and move together in time to music; however, participants used the batons in many different ways—to keep time to the beat, to communicate with one another, to hold above the head like cigarette lighters, and to "write" in the air.

For the above project, work by Wallace et al. (2013) has been important and inspiring. Wallace led a design inquiry into the experience of personhood in dementia focusing on a couple as a case study, where the wife (Gillian) had recently received a diagnosis of dementia. Wallace used design probes to create several pieces of digital jewelry for Gillian including a digital locket that displayed many pictures from Gillian's life when opened and a dress brooch (made of Gillian's old dresses) containing RFID tags and jewelry box that played back sound—often old songs or stories recorded by family—associated with the different fabrics in the dress brooch. In this paper, the authors reflected on the value of these digital objects and indeed the use of design probes, concluding that the creative nature of the probes allowed for expression while also "scaffolding" the process for the couple. More importantly, however, the design and wearing of this digital jewelry allowed Gillian to retain a sense of self and, it was hoped, also aided communication in further stages of dementia.

Thus we can see that—while assistive technologies are important and often integral to managing ageing and dementia—there are needs that may not be met by these technologies. While these needs may seem outside the scope of many technologies, creative methods of design can lead to effective, thoughtful, and useful technologies that can support a continued sense of self and psychological independence throughout dementia.

Photo-Sharing after Domestic Violence

The final set of studies that we will visit briefly in this chapter formed part of Rachel Clarke's research at Newcastle University (Clarke, Wright, Balaam, & McCarthy, 2013). Interested in participatory and community art and the place of both in designing with people facing challenging circumstances, Rachel entered the Angelou Centre, a black-led women's center based in the Newcastle upon Tyne. The center offers a range of "holistic women-only services for Black and Minority Ethnic (BME) women," and works at "strategic and national levels ensuring the voices of BME women are represented and heard."

An artist herself, Rachel initially carried out art sessions with the women frequenting the center but much as the SwaytheBand study above found, these women were mostly uninterested in art sessions. Instead, Rachel found a way "in" to their experience by participating in activities in which the women themselves were expert or which were directly relevant to their lives—which happened to be cooking. This approach allowed Rachel time to build trust with the participants, allowing them to gain confidence to come to sessions, and highlighted the importance of developing an approach that was sensitive to women's different situations.

In order to investigate these women's lives more deeply, Rachel extended her participative, ethnographic approach to carrying out digital portrait workshops. Women were provided with packs containing a digital camera, sound recorder, a portrait frame and a set of "inspiration tokens" based around people, experiences, objects and places. Workshops run over a series of weeks indicated that women appreciated working together in groups rather than working alone, that technical expertise in editing videos was problematic for the process, that researcher cooperation was necessitated, and that issues of anonymity and consent were prominent among participants.

The findings of the study, analyzed via narrative inquiry, yielded observations surrounding the nervousness of the women about making their stories public in some ways (i.e., visually), although they agreed to let their voices be heard in other ways—through narrative accounts of living with domestic violence or the collaboration of two of the women on creating a story. In one memorable interaction, two women, Saeeda and Zahrah, collaborated on a digital portrait that they wanted to focus on their friendship. The two documented a trip that they took together into the countryside where they had both purchased and worn the same dress. Their story highlights how the process of "making," for participants, can illuminate values that, while they can later translate to design, can also tell us much about the lives and experiences of our participants:

> The photographs here were used collectively to illustrate an ongoing commitment to what they felt was important to them, their friendship as a reflection of self. In addition to the written account, the photographs themselves highlight an embodied enactment of the women's similar experiences and their reciprocity through the act of photography as an illustration of the important value they placed on the commonalities within their friendship (Clarke et al., 2013, p.2531).

Rachel's and the participants' ultimate design was the Photo Parshiya (Clarke, Briggs, Light, & Wright, 2016)—a digital photo album that is held like a book and is portable or can be docked as a double touch screen tablet display on a bespoke crafted wooden base. Participants in this project made their own individually designed necklaces that reflect aspects of their identities; these necklaces are then linked with the tablet displays to create personalized photo collections.

Although the design itself is technologically innovative and aesthetically beautiful, and has been used to support workshops for women who attend the center, we think that the really interesting thing that emerged from the research in the Angelou Centre was the sense of design as an emerging process or a relationship between participants and researcher, where the sense of power (usually tipped in favor of the researcher, even in studies where participants are not characterized as "vulnerable") was balanced between multiple voices in the location to allow participants to act as experts in their own experience, and to share that experience with researchers.

More than that, the object itself and its perceived usefulness are not the important things that emerged from the research—instead, what is important is what the research and the design object make visible to us. In creating an object that allowed the careful curation of artworks and photographs to certain audiences, the research tells us that issues of anonymity are important to these participants, and have the potential to form a major stress in their life as survivors of domestic violence. This wish for anonymity is balanced with the wish to express themselves and live and cherish an open life, lived in the community of friends and family that they so value. The Photo Parshiya, therefore, is a useful object but also one that is emblematic of the findings of the study itself. This research, like much of the research that has come before in this chapter, is research through design, where the "making" itself construes a research method in and of itself, and what is important is not a set of objects but a set of ideas about how these participants experience life, and what role design might have within that life. We will discuss this form of research through design in the next section in more detail.

Themes in Designing with People Facing Challenging Circumstances

The above studies differ in terms of the population and settings that they work with—homeless people, people with dementia, and migrant women who have faced domestic violence—however, we hope that you can see similar threads that run through each of these studies in different ways. This penultimate section will discuss these themes as they relate to what we must keep in mind and put into practice when we carry out design research with people who face challenging circumstances.

Resilience, ability, and making choices

Earlier in this chapter we talked about disability, different abilities, and conceptualizing the development of these different abilities as adaptive responses to challenging circumstances. Resilience also plays significant roles in the design research mentioned above and in many of these cases it is linked tightly to creativity. In the Le Dantec and Edwards study, the difficult circumstances faced by the homeless people was tempered by their creative abilities and their resourcefulness that allowed them to live their lives in relative safety. The participant discussed above, who found a nontechnological way to manage his medication is one example; another in the same study used a noticeboard he had erected in his shelter to manage his appointments.

Part of this management extended beyond safety and health concerns to that of identity management—several participants in the study expressed fervent wishes to not "appear homeless": "I always find me someplace I can take a bath or take a shower or wash up. Because you know I like to keep clean, I'm always facing peoples, I didn't want to stand around in all dirty clothes." This identity management extended to online forms of self-management, where participants manipulated their MySpace profiles carefully and where one participant was very sure not to let his mother know that he was sleeping rough.

This careful manipulation of self-presentation to an outside world is something that is echoed in the studies by Jayne Wallace and Rachel Clarke; despite dementia patients often being characterized as disheveled or unconcerned / haphazard with dress and appearance, the digital jewelry worn by Gillian served to accent her appearance even as her dementia progressed, and beyond this, was a very deliberate decision to present aspects of herself to an outside world. In contrast, the women in the Angelou Centre study curated aspects of their selves and their histories for carefully selected audiences—sometimes just the researchers, sometimes just friends and coparticipants, and sometimes absolutely anyone *except* their husbands and their children.

In these ways, the participants in these above studies are not vulnerable populations whose daily lives are tumultuous and arduous; often, they don't even fit Woelfer's definition of susceptible. They are in fact resilient in the face of stress, able to make choices for themselves, and to echo Ungar's definition, above, capable of "navigating their way to the psychological, social, cultural, and physical resources that sustain their wellbeing." The key lesson in this is that this point of wellbeing, the resources that are sustaining for it, as well as the journey to that point itself, may not look like the way we imagine it—but we must respect it, and that is where participation becomes important.

Participation

The studies that we reference above are all broadly participatory in nature—the Le Dantec and Edwards study is a lot more like a conventional qualitative study, but the #Pat study picks up several points brought up in that and attempted to use participatory methods with a homeless population. Participatory methods are useful to use with populations we commonly perceive to be vulnerable because, when enacted sensitively, participatory methods can pave the way for compassionate design to take place; design that is empathic and which is as true as possible to our participants' lived experience.

Participation has been a value in many HCI projects for many years now; however, we need to examine the concept a little more closely when it comes to working with populations whose lived experience and ways of living and communicating are markedly different from our own. For people with dementia, verbal communication may wane, and although a carer can guide them through, for example, a cookery demonstration, they may not be truly participating or feeling a part of a larger activity or group. Conventional participatory design workshops can be problematic for people facing homelessness, as the above projects show, as their schedules and whereabouts are subject to change, often on a daily basis. And the migrant women in the Angelou Centre brought with them different cultural values to those of Rachel's, as well as troubled pasts that restricted the ways in which they felt comfortable expressing themselves. How can participation be possible for these people?

The answer (or at least our answer) is that it is relatively simple, and it lies in the turn to experience that we discussed earlier in this chapter. A turn to experience is critical in that it forces us to think about the multiplicity of perspectives from which our design or our research methods can be experienced. In doing this, we are able to recognize, affirm and prioritize the skills and abilities of our participants, rather than adapting for our own ends our research methods to take account of what we perceive as a vulnerability or a deficit.

A critical example lies in the ways in which issues of consent and communication were configured in the Swaytheband study. Institutional research ethics for people with dementia often demand not only the consent of the person with dementia (which is not held in particularly high regard, and is difficult to gain meaningfully) but the consent of a proxy, often a caregiver or occasionally a care institution. This is a position that potentially diminishes people with dementia; it positions them as nonpersons who are not capable of giving consent. Although the lack of capacity to give informed consent can be an issue in research, some studies circumvent this by seeking proxy consent to satisfy institutional ethics but also by seeking consent from the individuals themselves in a way that makes sense to them. For our research, we did this using picture cards, informal chats, verbal consent and ongoing consent, where the individuals are reminded multiple times throughout the research that they are a part of a research project but they may choose to leave at any time.

Beyond this, as we worked more and more closely with people with dementia, carrying out formal workshops and observing more informal creative sessions, we realized that we had been prioritizing one sort of participation above another, and that these participants were participating in ways that did not require sustained conversation or coordinated movements but instead communicated using short bursts of chat, song and poetry; movement, dance, waving, clapping, singing; eye contact, hand holding. This short description of an interaction with Valerie explains how some of these methods of communication played out: sitting and listening one day to an old record of Irish ballads, resident Valerie took the lead author's hand and the two swayed from side-to-side half-singing the songs. After a while this swaying turned into a dance as Valerie guided her hand in a sort of a twirling, twisting pattern in the air in rhythm to the music. The two paused after each song to applaud the singers, but Valerie reached for the lead author's hand immediately afterwards. At one point Valerie simply held her hand very tightly as they watched the television or chatted, or observed the people around them. Valerie squeezed her hand very tightly and ran her fingers over her knuckles. The lead author let her guide her hand again and she brought it very close to her face, rubbing it gently over her cheek and chin.

Our research progressed in a way that prioritized these forms of communication as participation and as participative in our design activities, which became something that unfolded with time more than it iterated from stage to stage.

A final point as it pertains to participation in these studies is the interrelationships between design, research and the lives of the researchers and the participants (who may become coresearchers). Frayling (1993) describes three sorts of design research—research into, through and for design. The research studies we have described as part of this chapter could be described as research through design—the design process itself is an inextricable part of the research process, and for many of these studies, the point is not to create a design but to learn through the design object as a communicator—for example, the SwaytheBand prototypes in the Morrissey and McCarthy research, above, act as conduits through which we learn about the role of movement, rhythm, embodiment, and communality in dementia care.

Bespoke design—does one size fit all?

This chapter began with a suggestion that design for all can sometimes end up as design for none (or at least, a limited amount of people). When we design, we must begin from a point where we are already designing with people facing difficult circumstances in mind; however, the value of bespoke design (which may be created *for and with* people perceived to be vulnerable but which should be valuable for people, full stop) should also be recognized.

In the above studies, the resulting designs were bespoke to varying degrees; the #Pat system and the Swaytheband systems could conceivably be used in multiple scenarios in different towns and different care homes (or homes in general) but Jayne Wallace's digital jewelry has a distinctively personal component, which was very much the critical point in her work—the aesthetic value of these designs, carefully crafted, beautiful to experience and to wear, would be lost if the idea were to be commercialized and rolled out as a set of recordable jewelry that can be bought for people suffering with dementia. Sometimes, the personal is the point.

Sometimes, the process is the point and the process is what is ultimately what is most rewarding in our research. Rachel Clarke's research in the Angelou Centre also resulted in a personal design—a sort of digital book-cum-picture frame, which played with ideas of anonymity and viewership by making some pictures visible to some people and keeping others for private viewing. Although the result was aesthetically designed and a beautiful object, nevertheless the process that she and her participants underwent was where most of the learning and the research took place. Design should not always be offered as a solution; sometimes it is a response to what has gone before, and is a part of an ongoing conversation rather than a stopper at the end.

The individuality of experience in all its richness within our participants means that occasionally, no matter how sensitive our collaborative processes and resulting designs are, our participants will not like them—they will shy from the design sessions and will reject the design object. Although this can be disheartening, it is to be expected. This is particularly true when designing with a community or a group in mind—for example, in our Swaytheband study we found the following:

- At the beginning, during the initial exploration of creative sessions, art sessions were not enjoyed by many residents, particularly men, one of whom (Ben) decried the activity as "girly."
- During field trials, some people disliked the busy music sessions and called for quiet, or disliked the music chosen (one woman, Gertie, would shout "crap! It's all crap!" at our singer when she sang a song she disliked), or did not wish to take part in the Swaytheband activity.

These individualities are often down to factors that are outside of the remit of the design sessions and are to be accepted—striving to make the activity something that is enriching for all could easily result in an erasure of what makes the design enriching for some. However, these "outsiders" to our design activities are as important as those who wish to participate—upon chatting to Gertie, above, I learned she preferred a certain kind of music, and felt a little bit "outside" of the rest of the group because of where she was sitting that particular day, both of which informed both our relationship and the design activities going forward.

Implications for HCI as We Study and Practise Today

The prior sections described how *resilience*, *participation* and *bespoke design/individuality in design* inform several design research studies that focus on people facing challenging circumstances. By paying attention to these themes and the ways in which they interact in our research, we can begin to build a view of a future designed by participants who themselves are able, participative, strong, and individual, whatever their circumstances.

But beyond these themes, carrying out design research with people facing difficult circumstances challenges several design ideas that apply for many people carrying out HCI research, particularly those working closely with participants and potential users.

Evaluation moving to the background

As we mentioned earlier, evaluation as a process in design is something that is necessarily moving to the background and becoming less important as it becomes more immediately important to understand prior experience (McCarthy & Wright, 2004; Petersen, Iversen, Krogh, & Ludvigsen, 2004).

Evaluation methods

Moreover, the methods we use to evaluate our designs and our technologies are changing; earlier techniques depended heavily on (often well validated) questionnaires and labor-intensive usability studies; however, how do you evaluate a technology the aim of which is to facilitate social reminiscence sessions for people with dementia, or emotional and creative expression by migrant women? A positive evaluation in these settings means the ability to enrich people's experiences or, else to allow resilience through learning and growth. How is this evaluated? Through documenting the process rather than intervening, observing, watching, potentially interviewing participants or using relaxed focus group methods. As our design methods change, so too must our evaluation methods.

Participation as a critical principle

Although participatory projects have a long history, the recent influx of studies purporting to be participatory indicates that HCI researchers still have a strong wish to include their participants in the design of their studies from as early a point as possible. However, we need to examine fully the ways in which these studies are participatory—taking points from McCarthy & Wright (2015). Who controls what is visible and invisible in these projects? To what extent can participants be said to be coauthors, coresearchers? Are their contributions equal (but not necessarily the same) as those of the researchers? If we are to call our projects participatory (and we should), we need to examine participation within them.

A reexamining of ethics

Designing for and with people commonly perceived to be vulnerable entails a lengthy ethical process that may see intense institutional review as well as navigating caregivers, external nonuniversity institutions, and other gatekeepers (Hugman, Pittaway, &

Bartolomei, 2011). A strong but nuanced sense of ethics is required at this stage; although institutional ethics boards may be satisfied with signed confirmation of informed consent, researchers must be sensitive to times when their participants may be uncomfortable with the research process as well as times when signed consent is not sufficient. Beyond this, the design object (if there is one) must also be thought of ethically—often, putting an object into a setting or asking people to live with an object or a service as an integral part of their lives can be construed as an intervention. If we are to truly value our participants as coresearchers, we must behave towards them in an ethical and enriching way.

Fidelity in qualitative methodology

Many of the above research methods necessitate a qualitative approach to data collection and analysis. However, many HCI researchers are not necessarily trained in using these methods, and we have noticed that many design research studies that have been published have been less than informative or transparent about their process of analysis. Although qualitative research requires a degree of interpretation by the researcher, it gains its validity through similar rigorous processes as those of quantitative analyses. We need to be open, honest, creative yet rigorous in the ways in which we utilize qualitative methodology if we want our research to be reliable, valid, and reflective of our participants' experience.

Summary

This chapter has explored what it means to design with people living in difficult circumstances. We set out to explore the word "vulnerable" and what it means to call an entire population "vulnerable"; eschewing the term, we then discussed how people who face challenges in their lives (whatever their source) often adapt to and overcome these challenges in particular ways. We considered different approaches to design and how we can best include the authentic experience of our participants in design. We then discussed five design research studies, exploring the experience of people living with homelessness, dementia and domestic violence, before analyzing three threads running through each of these studies—resilience, participation and bespoke design. We closed with future lessons for HCI research and practice.

References

Anastasiou, D., & Kauffman, J. M. (2011). A social constructionist approach to disability: Implications for special education. *Exceptional Children*, *77*(3), 367–384.

Anderson, I., & Christian, J. (2003). Causes of homelessness in the UK: A dynamic analysis. *Journal of Community and Applied Social Psychology*, *13*(2), 105–118.

Bavelier, D., Dye, M. W., & Hauser, P. C. (2006). Do deaf individuals see better? *Trends in Cognitive Sciences*, *10*(11), 512–518.

Bharucha, A. J., Anand, V., Forlizzi, J., Dew, M. A., Reynolds, C. F., Stevens, S., & Wactlar, H. (2009). Intelligent assistive technology applications to dementia care: current capabilities, limitations, and future challenges. *The American Journal of Geriatric Psychiatry*, *17*(2), 88–104.

Carrington, P., Hurst, A., & Kane, S. K. (2014). Wearables and chairables: inclusive design of mobile input and output techniques for power wheelchair users. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3103–3112). New York: ACM.

Cattaneo, Z., & Vecchi, T. (2011). *Blind vision.* Chicago, IL: MIT Press.

Clarke, R., Briggs, J., Light, A., & Wright, P. (2016). Situated encounters with socially engaged art in community-based design. In *Proceedings of the 2016 ACM Conference on Designing Interactive Systems* (pp. 521–532). doi:10.1145/2901790.2901882

Clarke, R., Wright, P., Balaam, M., & McCarthy, J. (2013). Digital portraits: photo sharing after domestic violence. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 2517–2526). doi:10.1145/2470654.2481348

Clarke, S. E., & Chenoweth, E. (2006). The politics of vulnerability: Constructing local performance regimes for homeland security. *Review of Policy Research*, *23*(1), 95–114.

Dewey, J. (1934). *Art as experience.* London: Penguin Books.

Frayling, C. (1993). Research in art and design. *Royal College of Art Research Papers*, *1*(1), 1–5.

Freedman, V. A., Kasper, J. D., Spillman, B. C., Agree, E. M., Mor, V., Wallace, R. B., & Wolf, D. A. (2014). Behavioral adaptation and late-life disability: A new spectrum for assessing public health impacts. *American Journal of Public Health*, *104*(2), 88–94.

Hugman, R., Pittaway, E., & Bartolomei, L. (2011). When "do no harm" is not enough: The ethics of research with refugees and other vulnerable groups. *British Journal of Social Work*, *41*(7), 1271–1287.

Le Dantec, C. A., & Edwards, W. K. (2008, April). Designs on dignity: perceptions of technology among the homeless. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 627–636). New York, NY: ACM.

McCarthy, J., & Wright. P. (2006). A dialogical approach to experience: Uncovering critical potential. *The Virtual*, 1–11.

McCarthy, J., & Wright, P. (2004). *Technology as experience.* Chicago, IL: MIT Press.

McCarthy, J., & Wright, P. (2015). *Taking [a]part.* Chicago, IL: MIT Press.

Morrissey, K. (2015). Community in dementia care and the role of familiar technology. In G. Avram, F. de Cindio & V. Pipek (Eds.), *International Reports on Socio- Informatics (IRSI), Proceedings of the Work-In-Progress Track of the 7th International ACM SIGCHI Conference on Communities and Technologies, 12*(1), 33–42.

Morrissey, K., & McCarthy, J. (2015). Creative and opportunistic use of everyday music technologies in a dementia care unit. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition* (pp. 295–298). New York, NY: ACM.

Petersen, M. G., Iversen, O. S., Krogh, P. G., & Ludvigsen, M. (2004). Aesthetic interaction: A pragmatist's aesthetics of interactive systems. In *Proceedings of the Fifth Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques* (pp. 269–276). doi:10.1145/1013115.1013153

Royal Commission on Long-Term Care. (1999). *With respect to old age: Long term care rights and responsibilities.* London: The Stationery Office.

Ryan, E. B., Bannister, K. A., & Anas, A. P. (2009). The dementia narrative: Writing to reclaim social identity. *Journal of Aging Studies*, *23*(3), 145–157.

Schroeder, D., & Gefenas, E. (2009). Vulnerability: Too vague and too broad? *Cambridge Quarterly of Healthcare Ethics*, *18*(02), 113–121.

Southern, J., Dillon, R., Potts, R., Morrell, D., Ferrario, M. A., Simm, W., & Whittle, J. (2013). Patchworks: citizen-led innovation for chaotic lives. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems* (pp. 2871–2872). doi:10.1145/2468356.2479554

Ungar, M. (2008). Resilience across cultures. *British Journal of Social Work*, *38*(2), 218–235.

Vines, J., McNaney, R., Lindsay, S., Wallace, J., & McCarthy, J. (2014). Special topic: Designing for and with vulnerable people. *Interactions*, *21*(1), 44–46.

Wallace, J., Wright, P., McCarthy, J., Green, D., Thomas, J. & Olivier, P. (2013). A design-led inquiry into personhood in dementia. In *Proceedings of ACM CHI 2013 Conference on Human Factors in Computing Systems* (pp. 2617–2626). New York: ACM Press.

Williams, V., Victor, C. R., & McCrindle, R. (2013). It is always on your mind: Experiences and perceptions of falling of older people and their carers and the potential of a mobile falls detection device. *Current Gerontology and Geriatrics Research*, *1*(1), 1–7.

Woelfer, J. P. (2014). Engaging homeless young people in HCI research. *Interactions*, *21*(1), 54–57.

Wright, P. McCarthy, J. (2010). *Experience-centered design*. Chicago, IL: MIT Press.

32

Innovative Accessible Interfaces

Carlos Duarte, David Costa, and Luís Carriço

Introduction

Designing interactive applications to provide a good user experience under different usage conditions is a difficult task—more so when the design space imposes requirements and constraints that are unfamiliar to most designers, leading to complex and apparently unsolvable problems. Designing accessible interfaces is one such design space. Accessibility refers to the need to address an individual's characteristics, be they endogenous, acquired, or situational. It imposes restrictions on the communication and interaction channels that are open between individual and device, but can also be understood as a search for a universal approach.

Looking for accessible solutions is not an easy task. Past research has explored different approaches to deal with it. These range from solutions that "fix" existing applications to ones that design novel services from scratch, be they tailor made for specific groups of individuals or created more broadly. Research has also explored methods for understanding needs and assessing solutions, as well as conceptual frameworks, tools, models, guidelines, and technology supporting the design and development of accessible interfaces.

This chapter addresses that research, summarizing the most relevant and recent contributions, mostly from a human-computer interaction perspective. It begins by introducing the accessible design space, focusing on its challenges, existing approaches for accessible design, including understanding the problem, designing a solution, and assessing its worth. It then addresses advances supported by recent technologies that are already contributing, or have a clear potential to contribute, to improving the accessibility of interactive systems.

The Accessible Design Space

The demand for accessibility in applications and interactive systems has a strong impact on the design space by imposing strong restrictions, on the one hand, and a need for wide coverage, on the other, which often results in conflicting requirements.

The Wiley Handbook of Human Computer Interaction, Volume 2, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

Anyone who has developed applications for blind people feels these restrictions. Most interactive applications are designed assuming the availability of visual perception. When this is absent or disrupted, the user experience is seriously compromised. Solutions that partially translate visual modalities to other interaction modes, like screen readers, are valuable, but often fall short if no application support is provided (Calvo, Iglesias, & Moreno, 2014). For example, while using a screen reader: the two-dimensionality of the visual presentation, whether tables, diagrams, hierarchies, is lost in the linearization of speech; huge amounts of content that are automatically discarded by a person using vision because they are not central to the task at hand (e.g. advertising, links and irrelevant sections), are scrupulously read by screen readers making any serendipity experience a nightmare; relevant images or emphasis are often not understood by technology, removing valuable content jeopardizing the user's comprehension.

The landscape for other disabilities, innate or acquired, is not better. For example, cognitive disorders require special care in the formulation and emphasis of text sentences (Rello, Baeza-Yates, Bott, & Saggion, 2013), in the design of sequences, and the consideration of unexpected interactions (Duarte et al., 2014), or in the centrality of images as means to convey content (Carmien & Fischer, 2008); motor disabilities need adequate spacing between interaction elements (W3C, 2008), or the concentration of interactive areas in small subregions, limited subsets of those reachable by other users (Guerreiro, Nicolau, Jorge, & Gonçalves, 2010); hearing impairments introduce challenges regarding subtitling and automatic transcriptions in near-real time (Kushalnagar, Lasecki, & Bigham, 2014), and regarding adequate alternatives to typical sound-based reminders, or the actual communication between deaf individuals and those without that disability (Paredes, Fonseca, Cabo, Pereira, & Fernandes, 2014).

Newell and Gregor (1999) analyze the demographics of disabilities, to emphasize the diversity of interactions that are required. An exhaustive classification of functions, deficiencies, and health dimensions can be found in the reference document of the World Health Organization (WHO, 2009). This includes a list of environmental factors that can cause disturbances in the function and capacity of all individuals in certain contexts. From both documents the various implications that these aspects have in the design of applications and systems that aim at accessibility are evident.

Moreover, impairments are often not isolated (WHO, 2009). For example, a loss of tactile sensitivity is common in some types of blindness; often, motor ability is compromised as a side effect of a cognitive disability. One of the most obvious causes for the confluence of multiple impairments is age (DESA UN, 2013). In some parts of the world this is such an important and expanding factor that the number of projects and research targeting elderly people has grown sharply in recent years (Vines, Pritchard, Wright, Olivier, & Brittain, 2015). To some extent, developing applications for the elderly could imply having to consider the constraints from all possible disabilities (Wagner, Hassanein, & Head, 2010). Of course, if one takes a fundamental ethical approach, then all interactive applications should address all impairments. Many (Stephanidis & Savidis, 1998; Vanderheiden, 1998) argue that this universal perspective is within the definition of accessibility.

This confluence of impairments requires considering more than one constraint in the design of applications, further challenging the solution space. But more than that, it introduces a first level of potentially conflicting constraints. For example, solutions that urge the proliferation of images, as a means of communication to bridge cognitive problems, can conflict with communication restrictions imposed by visual impairments. The use of speech-based interfaces to cope with vision impairments conflicts with hearing loss and the articulation of speech frequent in older adults (Coelho, Guerreiro, & Duarte, 2013).

Another factor is particularly striking regarding the configuration of the design space: the stigma caused by the use of solutions targeting a specific disability (Shinohara, 2012). It is common that applications designed for particular groups, with levels of usability/accessibility generally high within that group, are simply ignored because they are specific (Bichard, Coleman, & Langdon, 2007). The demand for accessibility should also take this factor into consideration, which introduces a second level of conflicting constraints: on the one hand the solutions must address the individual's specific needs emerging from his/her disability; on the other that specificity should not be evident, due to the social stigma it causes.

Designing Accessible Applications

Considering the challenges mentioned above, various researchers have discussed what are appropriate approaches to design accessible applications. We discuss here the aspects in which these approaches differ most from those approaches that do not consider accessibility. We will address the starting points and the endpoints of design, and we will present some of the most relevant methods for understanding requirements, the guidelines and design support techniques, the development tools and support technologies, and the recommended ways to evaluate the resulting solutions.

The pragmatic utopias

Stephanidis et al. (2012) identify two starting strategies for the design of accessible applications and systems: reactive and proactive. Regarding the starting points, the first involves the adaptation of existing applications and application environments (e.g. operating system) to the special needs of users. The authors include general assistive technologies (Cook & Polgar, 2007) as one of the results of this reactive strategy. These are hardware and software solutions for people with special needs, often orthogonal to the specific objectives of the applications. As such, they function as adapters that fit the systems and applications developed for common users to the qualities of users with special needs. This is the case with Braille keyboards or screen readers. Vanderheiden (1998) includes these technologies in the group of tools that special users can use in sometimes hostile environments. They do not necessarily result from the adaptation of existing applications but are instead instruments that adapt the environment to users.

Admitting that sometimes there is no other solution, Stephanidis et al. (2012) criticize the reactive strategy based on: (a) the opacity of some existing systems and applications that hinder the adaptation process; (b) the rapid technological evolution that always leaves adapted solutions behind the other versions; (c) the resulting lack of economic viability of the adaptation. As a result, the adaptations often suffer from a lack of quality. Others have also noted these criticisms. Edwards (1995) illustrates the difficulty of this adaptation with fitting blocks that in reality never fit exactly. One of the most obvious examples is the presentation of an image to a blind person. Although recognition techniques may enable the description of some image components for some illustrations (Moraes, Carberry, & McCoy, 2013), the fact is that, for some images, that can be virtually impossible. Nevertheless, the current success of reactive strategies is evident in the myriad of available application-agnostic assistive technologies (Cook & Polgar, 2007), some already included at the operating system level.

The proactive strategy, on the other hand, considers the special needs of users from the start of the design process. Interestingly, criticisms (b) and (c) mentioned above also apply to the creation of applications from scratch—probably even more so because design solutions for individuals with special needs are usually sidelined as a result of perceptions and prejudices of the cost-benefit ratio. That is why the defenders of this strategy usually adopt a conciliatory stand. Stephanidis et al. (2012) give several examples of the proactive use of general assistive technologies and Vanderheiden (1998) suggests that both approaches are "essential" to address the special needs of users efficiently. Again, a significant number of applications were designed from scratch considering special abilities, including some assistive technologies in the sense proposed by Cook & Polgar (2007).

Stephanidis et al. (2012) also include the dimension of scope in the definition of the proactive strategy. Their view is that the endpoint of design should aim to support the needs of all, which is the hallmark of approaches such as Design for All (Stephanidis & Savidis, 1998) and Universal Design (Vanderheiden, 1998). In its most fundamental versions, proponents of the proactive approach claim that applications should be developed on an agnostic core, from the interaction point of view, which adapts, automatically or not, to any user, platform, or metaphor, through appropriate interface manifestations for each need (Stephanidis et al., 2012). Important results of this strategy include rules for the development of accessible applications (Abascal & Nicolle, 2005); component libraries that comprise several alternative interaction modes (Stephanidis et al, 2012); agnostic adaptation of platforms on which originally inaccessible applications adapt to each individual (Biswas et al., 2013); or extensions to development environments that guide developers in their work (Stephanidis et al., 2012). Some of these results, in particular the one with perhaps greater disclosure at the political level, the Web accessibility guidelines (W3C, 2008) have, or potentiate for its comprehensiveness, a huge impact in the area of accessibility.

However, the design-for-all approach involves a possibly dangerous logic (Newell & Gregor, 2000), while trying to impose a goal that is, in most cases, utopian. Harper (2007) points out the difficulty, if not the impossibility, of finding an abstraction that is really agnostic to the interaction mode, or the omniscience required to meet all needs, wants, and enthusiasms of all potential users, not to mention, platforms, metaphors and others. Newell and Gregor (2000) interpose the user sensitive inclusive design as a way of restricting demands while keeping proactivity and the focus on inclusion. Vanderheiden (2000) considers the need to prioritize target groups and needs, taking into account the practical difficulty of a universal design.

The need for accessibility is an extensive design space, hampered by stringent limitations. These limitations require a focused effort, strongly centered on users (Harper, 2007) and their interaction capabilities (Newell, 1995). Usually it gives rise to applications designed for a specific group of users, either made from scratch, or adapted. The focus on coverage encourages the demand for universality (Stephanidis et al., 2012), and is usually only possible after observations arising from focused effort. When successful, however, the effect is much more notable.

Methodologies, methods, models, techniques, and tools

It is accepted that design approaches to accessible applications should be centered on users. User-centered design (UCD), in its original form (Norman & Draper, 1986), or in any of its many variants (Beyer & Holtzblatt, 1998; Nielsen, 1993), is therefore commonly adopted. This is not surprising given that the focus of digital accessibility is the abilities of individuals as they interact with technology.

Borsci, Kurosu, Federici, and Mele (2014) claim that the universal design or design-for-all approaches go further than general UCD on three fundamental points: the search for coverage; supporting theories; and the formalization of some of the methods used in the understanding process. To some extent, the first and the last are quite aligned as this formalization takes the form of rules that ultimately emphasize that the design target should aim at "the widest range of disabilities," or that agnostic cores should be at the center of technological solutions. Regarding theoretical support, though, the suggested differences are not exactly confined to accessibility issues per se. For example, the concept of psychotechnologies for socialization accommodates the social, social identity, psychological, and technological aspects of the understanding of the interplay between self, others, and the environment while interacting with the digital artifacts. These facets are deeply rooted in the notion of User Experience (UX) (Law, Roto, Hassenzahl, Vermeeren, & Kort, 2009). Moreover, the perspective is not at all alien to some UCD approaches. For example, contextual design (Beyer & Holtzblatt, 1998), raises explicit awareness of the understanding of the cultural, political, social, and psychological aspects in its methods, even if to a lesser extent.

Although the centrality of users is accepted, the truth is that often design falls short of expectations (Borsci, Kurosu, Federici, & Mele, 2014). While understanding requirements, the participation of users with special needs is minimal or even completely avoided. Analysts and researchers mostly opt for the use of preestablished design rules or base their design ideas on the characteristics of the target or disabilities. The reasons behind this gap apparently stem from two main difficulties: the interaction between researchers and people with special needs, and the difficulty of reaching a sufficient number of participants. The latter is sometimes hindered by protectionism, often justified, from formal and informal caretakers and institutions. It is usually a slow process of gaining the confidence of management structures, caregivers and participants, even before addressing the ethical or health aspects. The former commonly results from the analysts' and researchers' inability to deal with a specific disability, either by excess or lack of zeal, or simply by lack of a common communication channel. These questions are obviously more noticeable in the design of universal solutions. Finally, the main reason derives from the widespread perception that people with special needs are a minority with low income.

When users are finally involved, care should be taken to mitigate the difficulties of communication channels. For example, "thinking aloud" protocols are obviously inadequate for speech-impaired individuals. Conducting interviews with such individuals requires researchers who know sign language, or translators, which usually proves to be costly. Where disabilities are related to literacy, questionnaires are usually not possible. Regarding participatory design, Moffatt, McGrenere, Purves, and Klawe (2004) point out that the difficulties of communication between researchers and users with special needs are among the most important factors that affect the process. With regard to people with learning difficulties, Anthony, Prasad, Hurst, and Kuber (2012) recommend particular attention while planning and conducting participatory design sessions, which should consider different ways of thinking (visual versus textual) and users' personal context. Sahib, Stockman, Tombros, and Metatla (2013), in a study with blind users, mention the need for adjusting working methods so that communication is efficient. Again, the ways of thinking are stressed, as well as the forms of communication and interaction between individuals, and the difficulties inherent in the use of an exclusively verbal communication mode. Admittedly, even if accessibility were not an issue, participatory design methods should always take into account the characteristics of users. What is emphasized and recommended for accessibility cases is that the researchers should double their efforts as the differences and difficulties in communication are bigger.

Digging deeper into design, the recommendations for prototyping are similar. Prototypes should be adapted to the characteristics of target users, as expected. However, it is important to note that this adaptation should often not be limited to interaction modes under scrutiny. In some cases, the secondary aspects of a prototype become an obstacle for the study. For example, ignoring the physical aspect of a prototype and its tactile characteristics, because the focus of test is the collection and organization of available commands, can cause problems for blind users. In the case of some cognitive deficiencies, the risk that the user's attention focuses on the absent aspects of the prototype is very high. In general it can be said that, depending on the degree of disability, the prototypes used in design processes need to be rather closer to the targeted final products. As a consequence, several libraries and tools emerged targeting the creation of accessible prototypes (e.g., Kuber, Yu, & McAllister, 2007; Stephanidis et al., 2012).

From the systematic understanding of users and their activities, sometimes relevant abstractions in the form of interaction models emerge. In their most essential form these models expose the sensory and motor characteristics of individuals and their cognitive processes during interaction (Card, Newell, & Moran, 1983), as well as their psychosocial profiles (Borsci et al., 2014). Typically these models do not take a particular approach to the issue of accessibility, at least at this level. In other words, they consider the multiple characteristics of an individual, independently of whether special needs are involved. This follows Chi's (2013) line of argument regarding usability versus accessibility.

From a more pragmatic perspective, as computational representations of users, models considering accessibility are the foundation for tools (Stephanidis et al., 2012) or adaptive platforms (Biswas et al., 2013) that fit interaction with the abilities of users. Here, the specific characteristics of each disability are directly reflected in computational representations and guide the design, simulating the disability to aid the developer's understanding (Oikonomou, Votis, Tzovaras, & Korn, 2009), or enabling the developer to adapt the user interface on the fly (Costa & Duarte, 2011). Predictive performance models like GOMS (Card, Newell, & Moran, 1983) or KLM (John & Kieras, 1996) have been considered with regard to accessibility. Trewin et al. (2010) surveyed existing work and proposed a KLM for screen readers' users. There are also guidelines for design like those proposed by Abascal & Nicolle (2005) or the Web accessibility guidelines (W3C, 2008).

Finally, the task of evaluation resembles that of understanding requirements. The difficulties in accessing, involving and communicating with users remain (Borsci et al., 2014). Probably as a consequence, several researchers focused their work on automatic evaluation, with a particular emphasis on the Web (Leporini, Paternò, & Scorcia, 2006) (Fernandes & Carriço, 2012). Here, the existence of an extensively accepted set of recommendations (W3C, 2008) offers the baseline for automation. However, this has been criticized. Power, Freire, Petrie, and Swallow (2012) question the validity of WCAG itself, whereas Vigo, Brown, and Conway (2013), point out the limitations of automatic tools. Harper and Chen (2013) note the different evolution paces of technologies and recommendations thus rendering the evaluation process even more difficult. Nevertheless, there is a general agreement that WCAG and automatic evaluation have a particularly important role in the design process of accessible applications. This role is even more important when large-scale evaluations (Lopes e Carriço, 2010) or thorough rich applications assessment (Fernandes, Costa, Neves, Duarte, & Carriço, 2012) are targeted.

New Technologies for Accessibility

Wearable and body-based interaction

Recent technological advances and research projects about wearable technology and sensing devices have opened up new possibilities for using our bodies as interaction platforms. The always-present skin, when combined with our proprioceptive capabilities, plus the control we can exert over our limbs, are a sound alternative for interacting with personal devices, such as smartphones. This is an area with a great potential for improving the accessibility of interactive systems for multiple population groups, which is demonstrated by some recent work on this topic.

Current interactive devices mostly rely on visual mechanisms to guide the input and output cycle. Wearable and body-based interaction can provide nonvisual interaction mechanisms, or complement visual interaction mechanisms with other modalities (e.g. haptic). The visually impaired user group can greatly benefit from these advances, but other groups can also benefit. Motor-impaired users can also benefit by having the ability to perform input through gestures or body-based mechanisms more accessible than typing, and deaf users can benefit from vibration actuators, for example.

Skinput (Harrison, Tan, & Morris, 2010) was the first project to use skin as an interaction input system based on sensors. This system focuses on the forearm and hand parts of the body although authors state that it can be used with any other body part. Skinput uses the transverse waves created by the skin displacement from a finger impact and longitudinal waves emanated outwards from the bone towards the skin to identify the tapped location. To capture these signals, Skinput uses an armband composed by two arrays of highly tuned vibration sensors. The armband can be placed above or below the elbow but studies shown that placing the armband near the input area increases the accuracy of the results. The authors also tested the accuracy of the system in an eyes-free condition. Results show a 10% drop in accuracy compared with the other conditions. It was observed that the tapping radius error increased from 2 cm to up to 6 cm. The bioacoustic input modality is not strictly associated with any output modality: visual feedback projected on the forearm and wrist by using a pico-projector and nonvisual interface are some of the examples of the use of this input technique.

Point Upon Body (Lin et al., 2011) also appropriates the body as an input device. This approach uses an ultrasonic sensor attached to the wrist (e.g. watch or wristband). This system showed that the highest accuracy obtained was 84% for five points divided between the elbow and wrist. However, when adding more points the accuracy levels diminished about 20%. While the system implemented is not highly accurate, the authors previously conducted two important studies: one explored the division of the forearm (in an eyes-free manner) while the other showed the importance of the feedback from skin. The first study concluded that, typically, a person can divide his own forearm into six parts accurately (although few were able to divide it into seven and eight parts). While tapping their fingers on the forearm, participants did it instantly without taking time to be sure they tapped the right position. The authors concluded that users always show confidence in their actions although they may actually be wrong as the number of divisions increased. The studies showed that confidence and accuracy increased when closer to the end points (below elbow and wrist). Thus, these two points provide strong spatial hints for the users to strengthen their confidence.

More recent efforts (Makino, Sugiura, Ogata, & Inami, 2013) go beyond just touching and tapping actions, by using a new technique with photo reflective sensors, enabling pinching, pulling, twisting, and caressing. This system can detect one-dimensional tangential force in a cylindrical body area (e.g. the forearm). To detect when a user touches the arm, two armbands with photo reflective distance sensors are used. Deformation of the forearm is detected as a change in the skin surface's height. The photoreflective sensor emits infrared light to the skin surface and detects its reflected light intensity. Thus, by measuring the intensity, which depends on the distance to the skin, this technique measures the differences between the two sensors and detects the tangential force applied into the skin.

Dezfuli, Khalilbeigi, Huber, Muller, and Muhlhauser (2012) presented a palm-based imaginary interface to control the TV. First, the authors conducted an explorative study to observe how the participants would interact with their hand to perform a set of common commands with TVs, while at same time maintaining their attention on the TV screen. The authors report that all users (10 volunteers) used one hand as an input surface and the other hand's index finger was used as a pointer to the surface. Participants stated that these regions of their hand offered easy cues to interact without looking. The participants defined a total of nine landmarks on the palm's surface, which they believed to be easily touchable without any visual demand based on the proprioceptive sense. Participants stated that they would only map the most used functions of the remote control in their palms. As no information is displayed in their hands, users stated that the simpler the design the better. When asked about interacting with onscreen UI content, participants suggested 2D gestures such as swipe, scroll, and even drawing digits on the palm of their hand. They also suggested mapping the UI elements on the palm of their hands and by tapping in the corresponding position the element would be selected on the screen. By using an optical tracking system, a controlled experiment was conducted. The spatial precision of the touches was assessed and, on average, the diameter necessary to encompass 90% of all touches is 28 mm. The average effectiveness of the palm touches was around 96.8%. The finger landmarks registered worse results, with the pinky finger having the least effectiveness.

Wagner, Nancel, Gustafson, Huot, and Mackay (2013) describe a bodycentric design space that allows classifying and comparing multisurface interaction techniques. BodyScape reflects the relationship between users and their environment—specifically, how different body parts enhance or restrict movement within particular interaction techniques. BodyScape also focuses on the involvement of the user's body during interaction and the combination of "atomic" interaction techniques in order to manage the complexity of multisurface environments. In this work the authors compare two free-hand techniques, on-body touch and midair pointing, and a combination of both. The authors try to find which on-body targets are most efficient and acceptable and what performance tradeoffs are obtained by combining these techniques. Based on pilot studies, the authors defined 18 body target locations distributed across the body. The participants were asked to perform trials as quickly and accurately as possible. Results show that body targets located on the upper torso required less than 1400 ms to be touched, whereas targets on the dominant arm and on the lower body required more than 1600 ms. Targets on the dominant arm are touched more slowly than those on torso or shoulder. Participants were able to accurately touch on-body targets (92.4%) on the first try. Targets on the dominant arm were more prone to errors with participants selecting adjacent targets. Touching the lower part of the body registered the worse times but still only 200 ms slower. Participants' preferences for and within each body target were consistent with the performance measures. With both techniques combined, participants felt less comfortable to select targets in the dominant arm and non-dominant foot. The authors suggest that on-body targets should be placed on stable body parts (such as the upper torso) when tasks require precise and highly coordinated movements.

Mobile devices, touch surfaces, and Braille

Mobile devices have become an essential and indispensable communication tool in everyone's daily lives. Users of these devices have different characteristics and capabilities. However, mobile devices are not designed to be usable by everyone: they are visually demanding in both input and output channels and are shaped to a generic user model and then presented with a variety of adaptive mechanisms (Guerreiro, Jorge, & Gonçalves, 2010), some of which try to help users with impairments overcome the interaction barriers caused by the devices. Commercial solutions, both software and hardware, exist to this end. Google's TalkBack or Apple's VoiceOver are two successful examples of assistive technologies that have made a contribution by easing the access of blind users to Android and iOS devices. Recent research efforts are paving the way for novel solutions that might lead to increased accessibility for all users.

Kane, Bigham, and Wobbrock (2008) presented an alternative solution to the standard features offered by mobile device developers. The authors implemented Slide Rule, an interaction technique to improve the accessibility of multitouch screens when used by visually impaired users. This technique uses a set of four basic gestural interactions: one finger scan is used to browse lists; a second-finger can tap the selection; multidirectional flick gestures can perform additional actions, and an L-select gesture is used to browse hierarchical information. An evaluation of the system showed it to be faster for certain tasks but more prone to errors.

NavTouch (Guerreiro, Lagoa, Nicolau, Gonçalves, & Jorge, 2008) takes advantage of the user's capacity to perform a directional gesture and through it navigate the alphabet. This technique reduces the cognitive load required by the need to memorize the position where characters are located. Special actions like erasing characters, spaces, and other special characters are linked to the corner of the screen. The authors performed a study to evaluate and validate this approach for touch-screen mobile devices. The study suggests that, over time, users improved their navigation techniques, surpassing the standard MultiTap approach. In a later study, the words per minute (WPM) rate was compared between NavTouch and MultiTap, and results showed that NavTouch allowed participants to register higher WPM rates.

Oliveira, Guerreiro, Nicolau, Jorge, and Gonçalves (2011) assessed the advantages and disadvantages of four different techniques: QWERTY, MultiTap, which is the traditional method used in mobile devices with physical keypads, NavTouch (Guerreiro, Lagoa, Nicolau, Gonçalves, & Jorge, 2008), and BrailleType (Oliveira et al., 2011), which takes advantage of the capabilities of those who know Braille. The touch screen is divided into six large targets representing the dot positions. These targets are located at the corners and edges of the screen. Dots are marked by holding the target and double tapping on the screen on any area that accepts the word typed. Results showed that, despite QWERTY and MultiTap requiring searching for a specific character or group of characters along the screen, they allowed higher WPM rates than NavTouch and BrailleType. This is due to both these techniques requiring multiple gestures and inputs to access a character. When assessing the error rates, QWERTY and MultiTap were more prone to errors, while BrailleType was less prone to errors. Although participants experienced better results with both MultiTap and QWERTY methods, they stated that NavTouch and BrailleType were the easier methods.

Bonner, Brudvik, Abowd, and Edwards (2010) presented a novel solution for touch screen's accessibility issues when used by visually impaired people. No-look Notes is an eyes-free text-entry system that uses multitouch input and auditory feedback. It offers a two-step access to the 26 characters of the alphabet with a small number of simple gestures that remove the precise targeting required for example by QWERTY keyboards layouts. This system uses multitouch gestures to address the accuracy issues presented by other approaches; it makes use of the split-tap / second-finger tapping techniques to select the desired character. This enables single touch gestures for exploring the user interface. The UI of this method is arranged in an eight-segment pie menu. However, each segment of the pie contains a group of characters similar to the ones used on mobile phones with physical keypads. The users explore the pie menu by dragging their finger around the edges of the screen. When the users find the desired group of characters they select the group by using the aforementioned multitouch techniques. Then the user is presented with a new screen with the characters of the group. This time the characters are displayed alphabetically from top to bottom of the screen. The user selects the desired character by using the same input techniques. In addition, No-look Notes provides gestures for special characters and text editing. The authors performed a comparative study with VoiceOver. One of the first results observed is that this new system is not comfortable for the users as they mentioned that their hands and fingers were tired. When performance was compared, results show that No-look Notes outperforms VoiceOver. The majority of users stated that this novel approach was easier to learn and faster, and they felt in control of the approach.

Southern, Clawson, Frey, Abowd, and Romero (2012) proposed BrailleTouch, a six-key chord Braille keyboard for touch screens. This technique was designed to use the smartphone's screen faced away from the user and held by the two hands. The screen was divided into six parts in the same way as BrailleType, and other control keys such as space, backspace or enter were implemented through flick gestures. In the findings described by the authors, the performance achieved between 18 and 24 WPM although with 14.5% errors. Users, however, criticized the lack of a screen reader to confirm the characters typed.

Oakley and O'Modhrain (2005) developed a motion-based vibrotactile interface for mobile devices. The authors use three-axial acceleration sensing to directly control list position instead of using this sensor to control the rate of scrolling or directional movement. The goal is to link or associate certain specific orientations to determined list items. Two empirical studies were conducted to evaluate user performance and compare with the traditional use of accelerometers that control the rate instead of the positioning on the list. Results show that the higher the number of items in the list, the worse is the performance. However, workload didn't follow this trend. The authors also state that this new technique is considerably quicker to scroll to the adjacent item and to select an item of the list. In the second study, results show that users strongly preferred the position-based technique over the rate-based one. However, users commented that the latter technique lets them see the screen more easily. Based on both tests, the authors concluded that this method is only reliable for short lists that are not dynamically changing (e.g. address book).

Li, Dearman, and Truong (2010) present an interaction technique that leverages proprioception to access application shortcuts. Virtual Shelves uses proprioceptive capabilities to support eyes-free interaction by assigning spatial regions centered around the user's body to applications shortcuts. To measure the different angles between the body and the arm holding the cellphone the system uses an accelerometer and a gyroscope. A first study was conducted to measure the directional accuracy of visually impaired users. The results show that this type of users had significantly greater selection errors and selection time than sighted participants. But the authors are not sure if these results were influenced by the users' age (this study had users of around 45 years while the sighted study had users around 25 years old), which is known to be a degrading factor in proprioceptive capabilities. To account for these factors, the area of interaction was reduced from 7×4 regions to 5×3 (three rows of five icons each). In a second study the usability of this technique was assessed. The authors observed that users had more difficulties selecting the top areas. Also the middle column regions were more accurate (the center column). Overall 81.8% of the performed selections were correctly done. In a second phase of this study, the users were able to pick some applications and place them for selection on their more comfortable regions. The results show an increase of 6.5% on accuracy levels. All of the participants considered it to be faster than using their current interface. In this study, blind people presented social acceptance concerns regarding this technology, stating that they were afraid of hitting somebody with their arms when interacting in public.

Conclusions

Designing accessible systems and interfaces remains a challenge. However, novel technologies have the potential to increase the future of accessibility by exploring novel modalities that can complement, or even replace, current solutions. In this chapter, we provided an overview of the major challenges in this design space, and looked at recent work considering some of these modalities. In particular, we looked at wearable and body-based interaction that takes advantage of human proprioceptive abilities, and at mobile-based solutions exploring touch-based interactions, some of which are inspired by Braille knowledge targeting, in particular, the visually impaired.

While some of the presented research has not yet been applied specifically in accessibility related projects, its potential for this goal is undeniable. It is also comforting to acknowledge that, even if the focus hasn't been accessibility, various of the presented research efforts try to promote a systematic understanding of the users and the activities that are to be supported. BodyScape is an example of such, with the effort to characterize several interaction techniques through a body centric design space. Nevertheless, we should point out that body-based interaction alternatives, despite their accessibility enhancing potential, have still to embrace participation of users with special needs, and, in particular, define effective elicitation techniques for the accessible design process. In this regard, the work on mobile devices is more evolved at this point, as can be seen in several of the examples presented.

References

Abascal, J., & Nicolle, C. (2005). Moving towards inclusive design guidelines for socially and ethically aware HCI. *Interacting with Computers*, *17*(5), 484–505. doi:10.1016/j.intcom.2005.03.002

Anthony, L., Prasad, S., Hurst, A., & Kuber, R. (2012). A participatory design workshop on accessible apps and games with students with learning differences. In *Fourteenth International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)* (pp. 253–254). New York, NY: ACM. doi:10.1145/2384916.2384979

Beyer, H., & Holtzblatt, K. (1998). *Contextual design: Defining customer-centered systems (interactive technologies)*. San Francisco, CA: Academic Press.

Bichard, J.-A., Coleman, R., & Langdon, P. (2007). Does my stigma look big in this? Considering acceptability and desirability in the inclusive design of technology products. In C. Stephanidis (Ed.), *Fourth International Conference on Universal Access in Human Computer Interaction: Coping with Diversity (UAHCI)* (pp. 622–631). Berlin: Springer-Verlag.

Biswas, P., Langdon, P., Duarte, C., Coelho, J., Guerreiro, T., & Jung, C. (2013). An accessible, adaptive and multimodal digital TV framework and corresponding development tool. In *Science and Information Conference (SAI)* (pp. 28–37). Washington, DC: IEEE.

Bonner, M. N., Brudvik, J. T., Abowd, G. D., & Edwards, W. K. (2010). No-look notes: Accessible eyes-free multi-touch text entry. In *Eighth International Conference on Pervasive Computing (Pervasive)* (pp. 409– 426), Berlin: Springer Verlag.

Borsci, S., Kurosu, M., Federici, S., & Mele, M. L. (2014). *Computer systems experiences of users with and without disabilities.* Boca Raton, FL: Taylor & Francis.

Calvo, R., Iglesias, A., & Moreno, L. (2014). Accessibility barriers for users of screen readers in the Moodle learning content management system. *Universal Access in the Information Society*, *13*(3), 315–327. doi:10.1007/s10209-013-0314-3

Card, S. K., Newell, A., & Moran, T. P. (1983). *The psychology of human-computer interaction.* Hillsdale, NJ: Lawrence Erlbaum.

Carmien, S., & Fischer, G. (2008). Design, adoption, and assessment of a socio-technical environment supporting independence for persons with cognitive disabilities. In *SIGCHI*

Conference on Human Factors in Computing Systems (CHI) (pp. 597–606). New York, NY: ACM. doi:10.1145/1357054.1357151

Chi, E. H. (2013). The false dichotomy between accessibility and usability. In *Tenth International Cross-Disciplinary Conference on Web Accessibility (W4A)* (Article 19). New York, NY: ACM.

Coelho, J., Guerreiro, T., & Duarte, C. (2013). Designing TV interaction for the elderly—a case study of the design for all approach. In P. Biswas, C. Duarte, P. Langdon, L. Almeida, & C. Jung (Eds.), *A multimodal end-2-end approach to accessible computing* (pp. 49–69). London: Springer-Verlag. doi:10.1007/978-1-4471-5082-4 3

Cook, A. M., & Polgar, J. M. (2007). *Cook and Hussey's assistive technologies: Principles and practice* (3rd ed.). St Louis, MO: Mosby.

Costa, D., & Duarte, C. (2011). Adapting multimodal fission to user's abilities. In *Fourteenth International Conference on Human-Computer Interaction (HCII)* (pp. 347–356). Orlando, FL: Springer. doi:10.1007/978-3-642-21672-5 38

DESA UN. (2013). *World Population Ageing 2013.* United Nations, Department of Economic and Social Affairs, Population Division. New York, NY: DESA, Population Division, Population Estimates and Projections Section. Retrieved from http://www.un.org/en/development/desa/population/publications/pdf/ageing/WorldPopulationAgeingReport2013.pdf

Dezfuli, N., Khalilbeigi, M., Huber, J., Muller, F., & Muhlhauser, M. (2012). Palmrc: Imaginary palm-based remote control for eyes-free television interaction. In *Tenth European Conference on Interactive TV and Video (EuroiTV)* (pp. 27–34), New York, NY: ACM.

Duarte, C., Carriço, L., Costa, D., Costa, D., Falcão, A., & Tavares, L. (2014). Welcoming gesture recognition into autism therapy. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems* (pp. 1267–1272). New York, NY: ACM. doi:10.1145/2559206.2581337

Edwards, A. (1995). Computers and people with disabilities. In E. Alistair (Ed.), *Extraordinary human-computer interaction: Interfaces for users with disabilities* (pp. 19–44). New York, NY: Cambridge University.

Fernandes, N., & Carriço, L. (2012). A macroscopic web accessibility evaluation at different processing phases. In *International Cross-Disciplinary Conference on Web Accessibility (W4A)* (Article 18). New York, NY: ACM. doi:10.1145/2207016.2207025.

Fernandes, N., Costa, D., Neves, S., Duarte, C., & Carriço, L. (2012). Evaluating the accessibility of rich internet applications. In *International Cross-Disciplinary Conference on Web Accessibility (W4A)* (Article 13). New York, NY: ACM. doi:10.1145/2207016.2207019

Guerreiro, T., Jorge, J. A., & Gonçalves, D. (2010). Identifying the relevant individual attributes for a successful non-visual mobile experience. In *Twenty-eighth Annual Conference of the European Association of Cognitive Ergonomics* (pp. 27–30). doi:10.1145/1962300.1962309

Guerreiro, T., Lagoa, P., Nicolau, H., Gonçalves, D., & Jorge, J. A. (2008). From tapping to touching: Making touch screens accessible to blind users. *IEEE MultiMedia*, *15*(4), 48–50.

Guerreiro, T., Nicolau, H., Jorge, J., & Gonçalves, D. (2010). Towards accessible touch interfaces. In *Twelfth international ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)* (pp. 19–26). New York, NY: ACM. doi:10.1145/1878803.1878809

Harper, S. (2007). Is there design-for-all? *Universal Access in the Information Society*, *6*(1), 111–113.

Harper, S., & Chen, A. Q. (2013). Web accessibility guidelines. *World Wide Web*, *15*(1), 61–88. doi:10.1007/s11280-011-0130-8

Harrison, C., Tan, D., & Morris. D. (2010). Skinput: appropriating the body as an input surface. In *SIGCHI Conference on Human Factors in Computing Systems (CHI)* (pp. 453–462). New York, NY: ACM. doi:10.1145/1753326.1753394

John, B. E., & Kieras, D. E. (1996). The GOMS family of user interface analysis techniques: comparison and contrast. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *3*(4), 320–351. doi:10.1145/235833.236054

Kane, S. K., Bigham, J. P., & Wobbrock, J. O. (2008). Slide rule: Making mobile touch screens accessible to blind people using multi-touch interaction techniques. In *Tenth International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)* (pp. 73–80). New York, NY: ACM.

Kuber, R., Yu, W., & McAllister, G. (2007). Towards developing assistive haptic feedback for visually impaired internet users. In *SIGCHI Conference on Human Factors in Computing Systems (CHI)* (pp. 1525–1534). New York, NY: ACM. doi:10.1145/1240624.1240854

Kushalnagar, R. S., Lasecki, W. S., & Bigham, J. P. (2014). Accessibility evaluation of classroom captions. *ACM Transactions on Accessible Computing (TACCESS)*, *5*(3), Article 7. doi:10.1145/2543578

Law, E., Roto, V., Hassenzahl, M., Vermeeren, A., & Kort, J. (2009). Understanding, scoping and defining user experience: A survey approach. In *SIGCHI Conference on Human Factors in Computing Systems (CHI)* (pp. 719–728). New York, NY: ACM.

Leporini, B., Paternò, F., & Scorcia, A. (2006). Flexible tool support for accessibility evaluation. *Interacting with Computers*, *18*(5), 869–890. doi:10.1016/j.intcom.2006.03.001

Li, F. C. Y., Dearman, D., & Truong, K., N. (2010). Leveraging proprioception to make mobile phones more accessible to users with visual impairments. In *Twelfth International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)* (pp. 187–194), New York, NY: ACM.

Lin, S., Su, C., Cheng, K., Liang, R., Kuo, T., & Chen, B. (2011). Pub—point upon body: exploring eyes-free interaction and methods on an arm. In *Twenty-fourth Annual ACM Symposium on User Interface Software and Technology (UIST)* (pp. 481–488). New York, NY: ACM. doi:10.1145/2047196.2047259

Lopes, R., & Carriço, L. (2010). Macroscopic characterisations of Web accessibility. *New Review of Hypermedia and Multimedia*, *16*(3), 221–243. doi:10.1080/13614568.2010.534185

Makino, Y., Sugiura, Y., Ogata, M., & Inami, M. (2013). Tangential force sensing system on forearm. In *Fourth Augmented Human International Conference (AH)* (pp. 29–34). New York, NY: ACM. doi:10.1145/2459236.2459242

Moffatt, K., McGrenere, J., Purves, B., & Klawe, M. (2004). The participatory design of a sound and image enhanced daily planner for people with aphasia. In *SIGCHI Conference on Human Factors in Computing Systems (CHI)* (pp. 407–414). New York, NY: ACM. doi:10.1145/985692.985744

Moraes, P. S., Carberry, S., & McCoy, K. (2013). Providing access to the high-level content of line graphs from online popular media. In *Tenth International Cross-Disciplinary Conference on Web Accessibility (W4A)* (Article 11). New York, NY: ACM. doi:10.1145/2461121.2461123

Newell, A. (1995). Extra-ordinary human-computer interaction. In E. Alistair (Ed.), *Extraordinary Human-Computer Interaction: Interfaces for Users with Disabilities* (pp. 3–18). New York, NY: Cambridge University Press.

Newell, A., & Gregor, P. (1999). Extra-ordinary human-machine interaction: What can be learned from people with disabilities? *Cognition, Technology and Work*, *1*(2), 78–85.

Newell, A., & Gregor, P. (2000). User sensitive inclusive design—in search of a new paradigm. In *2000 Conference on Universal Usability* (pp. 39–44). New York, NY: ACM.

Nielsen, J. (1993). *Usability engineering*. San Francisco, CA: Morgan Kaufmann.

Norman, D., & Draper, S. W. (1986). *User centered system design: New perspectives on human-computer interaction*. Hillsdale, NJ: Lawrence. Erlbaum.

Oakley, I., & O'Modhrain, S. (2005). Tilt to scroll: Evaluating a motion based vibrotactile mobile interface. In *First Joint Eurohaptics Conference and Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems and World Haptics* (pp. 40–49). doi:10.1109/WHC.2005.138

Oikonomou, T., Votis, K., Tzovaras, D., & Korn, P. (2009). An open source tool for simulating a variety of vision impairments in developing swing applications. *Universal Access in Human-Computer Interaction. Addressing Diversity* (pp. 135–144). Berlin Heidelberg: Springer. doi:10.1007/978-3-642-02707-9 15

Oliveira, J., Guerreiro, T., Nicolau, H., Jorge, J., & Gonçalves, D. (2011). Blind people and mobile touch-based text-entry: Acknowledging the need for different flavors. In *Thirteenth International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)* (pp. 179–186), New York, NY: ACM.

Paredes, H., Fonseca, B., Cabo, M., Pereira, T., & Fernandes, F. (2014). SOSPhone: A mobile application for emergency calls. *Universal Access in the Information Society*, *13*(3), 277–290. doi:10.1007/s10209-013-0318-z

Power, C., Freire, A., Petrie, H., & Swallow, D. (2012). Guidelines are only half of the story: accessibility problems encountered by blind users on the web. In *SIGCHI Conference on Human Factors in Computing Systems (CHI)* (pp. 433–442). New York, NY. doi:10.1145/2207676.2207736

Rello, L., Baeza-Yates, R., Bott, S., & Saggion, H. (2013). Simplify or help? Text simplification strategies for people with dyslexia. In *Tenth International Cross-Disciplinary Conference on Web Accessibility (W4A)* (Article 15). New York, NY: ACM. doi:10.1145/2461121.2461126

Sahib, N. G., Stockman, T., Tombros, A., & Metatla, O. (2013). Participatory design with blind users: A scenario-based approach. In *Fourteenth IFIP TC 13 International Conference* (pp. 685–701). Berlin: Springer. doi:10.1007/978-3-642-40483-2 48

Shinohara, K. (2012). A new approach for the design of assistive technologies: design for social acceptance. *ACM SIGACCESS Accessibility and Computing* (pp. 45–48). doi:10.1145/2140446.2140456

Southern, C., Clawson, J., Frey, B., Abowd, G., & Romero, M. (2012). An evaluation of brailletouch: Mobile touchscreen text entry for the visually impaired. In *Fourteenth International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI)* (pp. 317–326), New York, NY: ACM.

Stephanidis, C., Antona, M., Savidis, A., Partarakis, N., Doulgeraki, K., & Leonidis, A. (2012). Design for all: Computer assisted design of user interface adaptation. In A. Savidis (Ed.), *Handbook of human factors and ergonomics* (pp. 1484–1508). New York, NY: John Wiley & Sons, Inc.

Stephanidis, C., & Savidis, A. (1998). Toward an information society for all: An international research and development agenda. *International Journal of Human-Computer Interaction*, *10*(2), 107–134.

Trewin, S., John, B. E., Richards, J., Swart, C., Brezin, J., Bellamy, R., & Thomas, J. (2010). Towards a tool for keystroke level modeling of skilled screen reading. In *Twelfth International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS)* (pp. 27–34). New York, NY: ACM. doi:10.1145/1878803.1878811

Vanderheiden, G. (1998). Universal design and assistive technology in communication and information technologies: alternatives or complements? *The Official Journal of RESNA*, *10*(1), 29–36. doi:10.1080/10400435.1998.10131958

Vanderheiden, G. (2000). Fundamental principles and priority setting for universal usability. In *Conference on Universal Usability (CUU)* (pp. 32–37). New York, NY: ACM. doi:10.1145/355460.355469

Vigo, M., Brown, J., & Conway, V. (2013). Benchmarking web accessibility evaluation tools: measuring the harm of sole reliance on automated tests. In *Tenth International Cross-Disciplinary Conference on Web Accessibility (W4A)*. New York, NY: ACM. doi:10.1145/2461121.2461124

Vines, J., Pritchard, G., Wright, P., Olivier, P., & Brittain. K. (2015). An age-old problem: Examining the discourses of ageing in HCI and strategies for future research. *ACM*

Transactions on Computer-Human Interaction (TOCHI), *22*(1), Article 2. doi: 10.1145/2696867

W3C. (2008). *Web content accessibility guidelines (WCAG) 2.0*. Retrieved from http://www.w3.org/TR/WCAG20/

Wagner, N., Hassanein, K., & Head, M. (2010). Computer use by older adults: A multi-disciplinary review. *Computers in Human Behavior*, *26*(5), 870–882. doi:10.1016/j.chb.2010.03.029

Wagner, J., Nancel, M., Gustafson, S. G., Huot, S., & Mackay, W. E. (2013). Body-centric design space for multi-surface interaction. In *SIGCHI Conference on Human Factors in Computing Systems (CHI)* (pp. 1299–1308). New York, NY: ACM.

WHO. (2009). *International Classification of Functioning, Disability and Health (ICF)*. Geneva: WHO Press.

33

A Concrete Example of Inclusive Design: Deaf-Oriented Accessibility

Claudia Savina Bianchini, Fabrizio Borgia, and Maria De Marsico

Introduction

Tremendous technological advances have allowed human-computer interaction (HCI) to be drastically improved. In particular, the expressiveness of interfaces to access and use digital services and resources has increased, while their difficulty for the final user has decreased over time. Communicating with a computer has evolved from using punched cards and command line interfaces to "natural interfaces." This can be compared with evolution from free climbing by ropes to the elevator and perhaps to the rocket to reach high targets. In fact, designing user interfaces (UIs) has developed into designing a more comprehensive user experience (UX). Notwithstanding this, and the apparent global access to all kinds of digital equipment and resources, research and practice in HCI still face many challenges. One of its most valuable achievements from the social point of view would be the full inclusion in the digital world of people with special needs. At present, we can still assess the presence of two main gaps giving raise to digital divide: a technological one and a "sociological" one. The former refers to the difficulty of accessing some enabling technology, for example, wide bandwidth channels, which can still be observed in some settings. The latter refers to the difficulties still experienced by specific categories of users. First of all, it is worth reminding the distinction by Prensky (2001) between "natives and digital immigrants", that still holds in many social groups, and refers to the difficulty of adapting to new technologies by those who were born and grew before their introduction. A further affected category includes people experiencing some kind of limitation in exploiting traditional human communication channels. This not only includes disabled people but also older users whose perceptual abilities may have been degraded by age.

Since the late 1990s, lists of best practices to facilitate access to the Internet have been developed and formalized. The main attempts to address general accessibility issues have been made by the World Wide Web Consortium (W3C), with the Web Content Accessibility Guidelines (WCAG) document. This document aims to

The Wiley Handbook of Human Computer Interaction, Volume 2, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

anticipate the needs of people with disabilities, to enable them to enjoy, freely and independently, the contents of Internet. In fact, WCAG2.0 states in its own abstract that "following these guidelines will make content accessible to a wider range of people with disabilities, including blindness and low vision, deafness and hearing loss, learning disabilities, cognitive limitations, limited movement, speech disabilities, photosensitivity and combinations of these" (Caldwell, Cooper, Guarino Reid, & Vanderheiden, 2008).

Blind people and their needs are the group that mostly appears in research addressing accessibility. However, limitations suffered by the deaf are often underestimated and therefore scarcely tackled. (Despite what some people— not working closely with the deaf—may think, there is nothing offensive in the word "deaf" when used to speak of the people belonging to the linguistic and cultural minority called the "deaf community"; for this reason, we have decided to use this term throughout the chapter.) This often depends on a kind of ignorance or misunderstanding of the real nature of their communication difficulties. Even in recent years, many HCI researchers still thought that it was sufficient to substitute audio with text, or to use subtitles in videos, to address usability needs of deaf people. We will point out how this caused a lack of attention with respect to very specific problems, which must be addressed in a very specific way. In many cases changing the channel is not sufficient, but it is also necessary to rethink the structure of the information (see for example the e-learning platform designed expressly for deaf people in Bottoni et al., 2013). The digital world offers unique occasions for integration but people with special needs risk losing them. This risk is especially concrete when such needs are scarcely appreciated, as happens for deaf people. This chapter aims to increase the awareness of deaf problems in the digital world and to propose a comprehensive solution for their better inclusion. We will therefore start by providing a short history of the evolution of the understanding of deafness issues, and of strategies to address them. Afterwards, we will briefly present the design and evaluation of the first nucleus of SWORD (SignWriting Oriented Resources for the Deaf), a comprehensive digital framework whose final goal is to support deaf people in exploiting the opportunities of the digital world.

Vocal Languages, Sign Languages and Accessibility

Despite deafness appearing in the second place in the list of disabilities considered by the WCAG and mentioned above, attention to deaf-specific problems has been late. The 1999 WCAG1.0 guidelines (https://www.w3.org/TR/WCAG10/) deal mostly with labeling and transcription of audio content, leaving out alternatives related to sign languages (SLs). The only mention relates to the possible management of videos. In practice, the suggestion is to complement the audio content with subtitles and, if possible, with a translation into SL. Of course, these guidelines are dictated by a sincere effort to solve the accessibility problems of the deaf. However, they partly reveal the misunderstanding by which it is assumed that a written transcription of audio content is enough to support deaf users. As we will emphasize, text is strictly connected to verbal languages, and it relies on underlying linguistic structures that are different from those developed by deaf people. Therefore, it presents several difficulties for them even in everyday communication. The 2008 WCAG2.0 guidelines (https://www.w3.org/TR/WCAG20/) represent a progressively increasing awareness of deaf needs and better addresses such issues. For instance, success criterion 1.2.6, whose satisfaction is necessary to get the highest level of compliance (AAA), states that "Sign language interpretation is provided for all prerecorded audio content in the form of synchronous media types." This opens the debate about at what extent videos or related techniques can really solve all accessibility problems encountered by deaf people.

In any case, WCAG guidelines are insufficient to grant the navigating autonomy that they aim at achieving. This is explicitly stated in the document:

> Note that even content that conforms at the highest level (AAA) will not be accessible to individuals with all types, degrees, or combinations of disability [...] Authors are encouraged [...] to seek relevant advice about current best practice to ensure that Web content is accessible, as far as possible, to this community (WCAG2.0, 2008)

and it is in the quest for these best practices that this chapter will be developed.

A starting factor is the heterogeneity of the population with hearing problems, which can be considered from different points of view: medical, sociocultural, and linguistic. In the medical perspective, two distinct criteria hold, namely degree and age of onset of deafness. Under the sociocultural perspective, it is important to consider the closeness to the "deaf community" and self-identification as "Deaf" or "hearing impaired." Last but not least, the linguistic point of view considers mastery of a vocal language (VL) and/or of a SL. Among these pillars there are extreme cases. At one extreme, we find slightly deaf elderly, who communicate in VL, and consider themselves hearing people yet lacking some hearing. At the other extreme, we find prelingual deaf—those born or turned deaf before learning to speak—who communicate in SL and feel Deaf. In the middle, we have to consider many shades, such as deep prelingual deaf who do not know SL, or mild deaf who use SL. One common element among most of the prelingual deaf is, indeed, the problematic relationship with the VL, and in particular with its written form. A report drawn up in France (Gillot, 1998) highlighted how 80% of the early deaf are considered almost illiterate, despite having attended school; this percentage applies to both signer (using SL) and oralist (using exclusively VL) early deaf. In a nutshell (for more details the interested reader can refer to Goldin-Meadow, 2001), the problem arises from a lack of exposure of the deaf child to the VL since early age. This creates a gap that, regardless of the possible exposure to the SL, determines a patchy acquisition of the VL. Given the strict cognitive/linguistic relationship, such shortages are then reflected in both the production and interpretation of written text. In fact, deaf people tend to reflect their visual organization of the world over the organization of language. Given this peculiarity, they find significant difficulties in both learning and mastering VLs, even in their written form (Perfetti & Sandak, 2000). Moreover, even if deaf people succeed in mastering a VL, dealing with VL content may prove quite a tiring task for them, unless it is performed for a short time. In fact, it can be observed that most of them prefer to communicate using SLs (Antinoro Pizzuto, Bianchini, Capuano, Gianfreda, & Rossini, 2010). Dealing with the issues of deaf-oriented accessibility using only written VL is therefore quite unrealistic (Borgia, Bianchini, & De Marsico, 2014), and any VL-based solution (VL captioning and transcription) to overcome the digital divide rarely solves the problem completely. In other words, captioning-based accessibility design may support the needs of people who become deaf after the acquisition of speech and language (postlingual deafness). However, issues related to prelingual deafness are addressed seldom and poorly. Deaf illiteracy becomes even more relevant for our purpose given the importance of textual information in most applications, and of written communication in the global network. In fact, despite the regular increase of multimedia content, text remains the main mode of content diffusion over the Internet; even on sites like YouTube, which are especially oriented toward video contents, we can observe that titles, descriptions, comments, and almost all the search tools are mainly based on text.

Summarizing the above considerations, we can suggest that the heterogeneity of the deaf population must be reflected in a variety of choices for accessibility: subtitles for the “late” deaf, and alternative solutions for the prelingual or early deaf, whether signers or oralists. It is important to take into account the preference for SL by the signers, and the difficulties encountered by the oralists with the VL, despite this being their only language. Unfortunately, to the best of our knowledge, no universally accepted instrument exists to profile a deaf individual on all of the factors that we mention. However, some interesting examples can be found regarding specific application fields, such as e-learning (Salomoni, Mirri, Ferretti, & Roccetti, 2007).

Although we use the singular expression “sign language,” we are referring to a class of languages rather than a unique one. Sad to say, as with VLs, even SLs present national and regional differences in the use of signs. However, given the strong perceptual basis of such languages, people signing differently can still understand each other better and more quickly than those speaking different VLs. Indeed, the latter usually do not understand each other at all, and tend to use gestures to help communication. In this work we will be focusing more closely on the deaf signers’ case. First, we will show the peculiarity of SL, the problems arising from the lack of a (universally accepted) written form for it, and why signed videos are not enough to allow for an accessibility comparable to that offered by the script. We will then present different systems created for the graphical representation of the SLs and, finally, we will propose our solution. It entails the use of two software tools designed to be part of a more comprehensive framework.

A Close View of Sign Languages

A SL is a language that uses the visual-gestural channel to convey the full range of meanings that, in VLs, are expressed through the acoustic-vocal channel. As stated above, it is not the language used by all of the deaf. However, the fact that SL exploits the intact visual channel, and related cognitive and communicative structures, makes it more accessible to them. Failure to attain a more widespread diffusion of SL to the prelingual or early deaf is mostly due to historical and cultural factors (a brief history of deaf education is provided by Stokoe, 1960). In fact, since antiquity, philosophers have questioned the link between thought and speech, often arguing that the absence of speech corresponds to the absence of mind (hence the recent phrase “deaf and dumb,” in which “dumb” means both “not talking” and “stupid”). Starting with the Renaissance, some tutors began to think about strategies for deaf education. Two schools of thought arose: the one formerly most diffused, namely the “oralist” school, concentrates its efforts on teaching the spoken VL at all costs, training a deaf person to communicate through lipreading and speech, and often at the expense of the content; the second school, the "signing" one, favors taking advantage of the SL for communication for the the full provision of teaching contents, and often neglects teaching of spoken VLs. Of course, the latter is based on the observation that deaf naturally use a kind of visual-gestural communication system, and on the purpose to exploit it also for education. In the second half of the 18th century, the Abbot de l'Épée founded in Paris the first school for the deaf. It developed a teaching method based on the use of the SL to replace the oral communication, and of written French to interface with the rest of the society.

This revolution in deaf education, however, was accompanied by severe criticism from those believing that deaf children should learn to speak, and not just to write. After a century and a half of wrangling, the Milan Congress on Deaf Education sanctioned, in 1880, a ban on SL in the education of deaf people, trumpeting the slogan "signs kill speech." The consequences of those decisions were more or less severe depending on the country, but the entire field of deaf education was touched. For instance, in France, home of the Abbot de l'Épée, it was necessary to wait until 1977 to have the ban removed, until 1984 for the establishment of a bilingual school for just a few deaf students, and until 2005 to have a law on the accessibility of handicapped people, sanctioning the right to use SL. In the United States, although unaffected by the Congress of Milan and considered a pioneering country in this matter, SL is recognized at state level as a "foreign language" or as a "language for education" at best. Discussions between pro- and anti-SL are still on the agenda, and are particularly virulent in contexts without laws about SL recognition (like in Italy). In these cases, many lobbies try to convince the legislature that the recourse to SL prevents VL mastery.

These educational discussions pair with the linguistic diatribes. It was necessary to wait until the 1960s, and for the works of the American linguist William Stokoe (1960), to obtain the first analysis of the linguistic structure of the SL. Those studies finally let it into the Olympus of "real languages," like the many VLs. The idea of Stokoe and his successors was based on the search for common features between SL and VL. This approach made it possible to show the linguistic features of SL. However, it was limited to a very small number of signs with linguistic characteristics strictly similar to VL: they were *sign words*, i.e. isolated signs that may be listed in SL vocabularies and easily translated with a single VL word. A further limit of these works was that they were focused on the manual components of SL, and this limited the exploration of the full expressive power of the language. Starting in the 1980s, Christian Cuxac and his associates, after the implicit assumption that SL belongs to the "real languages," developed a different model—the so-called "semiotic model" (Cuxac, 2000; Cuxac & Sallandre, 2007). Researchers taking this line aimed to show that SL is to be considered as a true language, with a richness and expressiveness comparable to VLs, despite presenting characteristics fundamentally different from them. Differences are due to the use of the visual-gestural channel rather than the acoustic-vocal one. The works by Cuxac (1996, 2000) on iconicity in LSF (French Sign Language) focused attention on the nonmanual components of signs, which are of paramount importance in many linguistic constructions of SL. Cuxac demonstrated that information in SL is also transmitted through eye gaze, facial expressions, and body movements. This research then spurred analogous investigations on other SLs, such as Antinoro Pizzuto's research on Italian Sign Language (LIS) (Antinoro Pizzuto, 2008; Antinoro Pizzuto, Chiari, & Rossini, 2010). According to the semiotic approach, SL has two ways of expressing concepts, whatever the national variant (Cuxac & Antinoro Pizzuto, 2010) is:

- the “lexematic units” (LUs), i.e. the *sign words* analyzed by Stokoe, which can be easily translated using a single VL word; LUs are mostly characterized by manual production components (configuration, orientation, location and movement of the sole hands);
- the “highly iconic structures” (HISs), i.e., complex structures with an underlying highly iconic content, which cannot be translated with a single VL word; in general, a HIS corresponds to several words, sometimes even to whole SL sentences (see Figure 33.1), and are produced by bringing into play both manual and nonmanual components, e.g., hand position, facial expression, gaze direction, etc.

Some researchers still dispute the linguistic value of the HISs, considering them a mere *mimic gestural complement* to the LUs. However, it should be stressed that, although almost all the isolated signs in the SL dictionaries are LUs, several researchers (including Cuxac & Sallandre, 2007; Antinoro Pizzuto, Rossini, Sallandre, & Wilkinson, 2008) have shown that 95% of the signs used in SL narratives are HISs (and not LUs), making it possible to firmly challenge any stance doubting their linguistic character (unless one accepts the existence of a language of which 95% is nonlinguistic).

The multiplicity of articulators that come into play in the production of HISs (the two hands, the body, the facial expression, the eyes, etc.) makes SL a language characterized by multilinearity and simultaneity. In practice, VL produces 1D expressions that either develop along time (spoken), or along space (written). Differently from VL, SL expressions develop along 4 dimensions: 3D space and time. These characteristics make it impossible to carry out the “literal” translation *sign-to-word* from SL to VL. It should also be pointed out that even the syntactic structure of a national SL does not match that of the local VL: SL has a “visual” grammar. For instance, the sentence “John gives a flower to Mary” will be signed by placing the sign for John in

(a) (b)

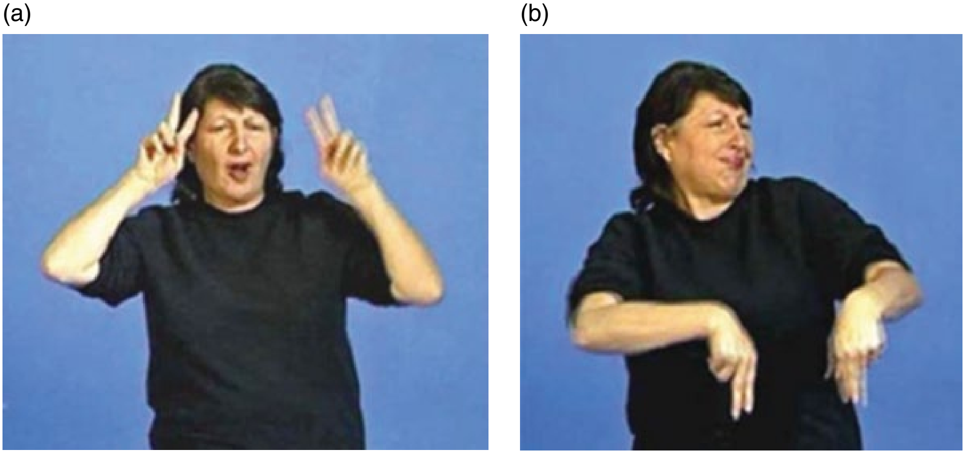


Figure 33.1 (a) French Sign Language from Corpus “LS Colin” (Cuxac et al., 2002): Lexematic Unit [HORSE]; (b) Highly Iconic Structure of a horse galloping. *Source:* Garcia and Sallandre (2014) Figure 11, p. 249. Used with permission.

a space A, the sign for Mary in a space B and making the sign for "to present" (which may assume a different manual form for a flower or a ball) move from A to B.

A final key feature of SL is that it is a language that has not developed a writing system accepted by a large part of either the national or the worldwide deaf community. This trait is not rare, as only 10% of the world's languages (spoken, however, by 90% of the world's population—Lewis, Simons, & Fennig, 2015) have their own writing system. The fundamental difference with the other oral-only languages is that all the VLs can be represented by the International Phonetic Alphabet (IPA), which does proceed through the acoustic-vocal channel. This is impossible for SL, due to the fact that it is completely detached from this channel. In other words, the specificity of SLs does not allow an equivalent solution, due to the lack of a kind of atomic articulatory element with the same role of spoken letters. Furthermore, the characteristics of multilinearity, simultaneity, and iconicity make it difficult, if not impossible, to adapt to the SL, a system based on IPA. In the remaining part of this chapter, we will outline the different systems that, despite these difficulties, have tried to overcome the problem of the SL graphical representation. To describe these systems it is necessary to introduce and explain beforehand some linguistic terms, which are not usually used with the same meaning as in other fields, i.e. "transcription" and "writing."

- "Transcription" is the use of a graphics system to represent a linguistic phenomenon present in a spoken or signed production; IPA can be used to represent the phonology of a language, allowing every detail of a phonological production to be caught and analyzed; in the same way, a graphical system for SL should catch each movement that the signer does with the intent to convey a meaning; it is precisely to achieve this aim that the systems we discuss below have been developed.
- "Writing" has many valid meanings but the one we will take into account is the action that allows language to be produced by means of a graphic representation of the language; contrary to the transcription, writing is a direct linguistic realization and not the graphical transposition of a previous production; none of the systems that we will examine, except SignWriting (Bianchini, 2012; Sutton, 1995)—which underlies the software to be discussed—was conceived to write the SL.

Not only is there a lack of a writing system for SL; there is also a lack of a suitable transcription instrument. It must be emphasized that, for linguistic researchers, the graphical representation of SL is a real challenge, going well beyond a problem of accessibility: in fact, solving this problem would be the first step towards new frontiers for the scientific research on the language structures, with significant implications both for "pure" linguistics and for pedagogy. The main goal of the tools that we will discuss is therefore to allow deaf users to find and use the language they prefer in interfaces as well as digital resources. In addition, a side effect of producing effective tools would be to facilitate linguistic research too.

Examples of Strategies to Represent / Communicate Digital Resources for the Deaf

As discussed above, SLs have been long considered as a purely mimic form of communication. As a consequence, their full status as languages has long been questioned as well, when not being completely denied. This heavy historical heritage has long conditioned research and design in a twofold way: by hindering the comprehension of the real problems that deaf find with vocal languages, and by delaying the search for effective communication strategies to support them. This affects both the education process and, more recently, the access to the digital world of information. The lack of awareness of the deep structural (cognitive) differences between VLs and SLs has also nurtured the illusion that it is sufficient to fill a purely sensorial gap by simply substituting auditory information by written text, as mentioned above when discussing WCAG guidelines. As emphasized above, the leitmotif of most accessibility guidelines that should address the needs of deaf people, is limited to textual captioning and audio content transcription (Fajardo, Parra, Cañas, Abascal, & Lopez, 2008). This practice implies providing textual (VL) captioning wherever a resource on the system (a website, an application, etc.), either audio or video clip, exploits the audio channel. It should be evident at this point that this is not a solution because both signing and nonsigning people (those subject to oralist techniques) make similar mistakes with spoken and written language (Antinoro Pizzuto, Bianchini et al., 2010). On the other hand, despite the advances in the linguistic comprehension of their structure, SLs have not still developed a writing system of their own. This is the main reason for a special kind of digital divide involving deaf people that still seriously limits the possibility of providing information, for example, electronically and on the Web, in a form "equivalent," from the cognitive point of view, to the signed content. Think for example of the difference between a static text and a search: while the former can be substituted by a clip in SL, how is it possible to support the latter?

As a consequence, a deaf person finds in the digital world the same numerous and difficult barriers faced in everyday life. An example is given by the impossibility of exploiting the "information scent" (Chi, Pirolli, Chen, & Pitkow, 2001), while surfing through Internet. During a search, each user exploits "semantic" traces related to the desired information to judge which link to follow or which site to browse. The so called "trigger words," i.e., words used for links and menu items, are a fundamental element in this process. From the point of view of deaf people, these semantic traces are generally available in a language that is not their "native" or preferred language, and in which they often have insufficient reading proficiency; we can experience a similar difficulty when browsing a site in a foreign language that we may know, but whose more subtle aspects are fully grasped only by a native speaker. As a matter of fact, some studies (Fajardo, Arfe, Benedetti, & Altoé, 2008) have investigated and reported the difficulties that deaf users find in gathering the information they need through textual traces. The gradually increasing awareness of the research community about deaf difficulties is producing the first results. Accessibility issues have been mostly tackled in connection with more "serious" disabilities like blindness but now they are taken into account also in relation with deafness. As examples of the new interest raised by related problems, we can mention the special issue "Emerging Technologies for Deaf Accessibility in the Information Society" of the international journal *Universal Access in the Information Society*, published in February 2008 (Efthimiou, Fotinea, & Glauert, 2008), and the frequent presence of related papers in human-computer interaction (HCI) conferences. Related research often deals with facial / gesture analysis for SL recognition from signed video (a review can be found in Ong and Ranganath, 2005), or with sign synthesis and sign language animation, possibly through avatars (Elliot, Glauert, Kennaway, Marshall, & Safar, 2008). It is worth underlining that most research in automatic analysis of SL has focused on recognizing the lexical form of sign gestures in continuous signing, aiming at scaling well to large vocabularies. However, "successful recognition of lexical signs is not sufficient for a full understanding of sign language communication. Nonmanual signals and grammatical processes that result in systematic variations in sign appearance are integral aspects of this communication but have received comparatively little attention in the literature" (Ong & Ranganath, 2005). It is clear that this pessimistic observation refers to the goal pursued by most computer scientists, who approach this topic with the aim of building a kind of automatic interface, able to work as an interpreter on behalf of deaf people. However, most efforts along this direction have to face a number of technological as well as conceptual problems. A full direct translation between vocal languages and sign languages must overcome the problem discussed above of mapping a 1D (verbal) flow onto a 4D (signed in space and time) flow and vice versa. Automatic interpretation of SLs, and especially a bidirectional one, is a goal that is still hard to achieve, if not unfeasible:

> For deaf persons to have ready access to information and communication technologies (ICTs), the latter must be usable in SL—it must include interlanguage interfaces. Such applications will be accepted by deaf users if they are reliable and respectful of SL specificities—use of space and iconicity as the structuring principles of the language. Before developing ICT applications, it is necessary to model these features, both to enable analysis of SL videos and to generate SL messages by means of signing avatars (Braffort & Dalle, 2008).

We can conclude that automatic interpretation as well as "translation" is quite a chimera at present.

Less ambitious, but possibly more effective, proposals in the literature entail using recorded videos without any automatic interpretation attempt. Overall, despite the efforts made by the W3C (Caldwell et al., 2008), widespread support for SL in the digital world is still far from being realized. Among the few exceptions, it is possible to mention the Dicta-Sign project (Efthimiou et al., 2010). Videos, more specifically signed videos, are typically the most widespread technique for SL inclusion. Concretely, signed videos are video clips representing one or more people producing signs. In this regard, some research projects have developed a series of techniques to allow deaf people to access digital information through different forms of deaf-oriented hyperlinking (Fajardo, Vigo, Salmerón, 2009). A seemingly simple idea is the basis of the Cogniweb project (Fajardo, Parra et al., 2008), which proposes two different techniques to equip a Web page with SL videos intended to support navigation. In the first proposed technique, a video frame is located at the bottom of the page, and starts the corresponding SL sequence as the user moves the mouse over a link. In the second technique, a mouseover-activated signed video is included within each hyperlink. Two tests demonstrated that deaf people can navigate more efficiently using the second technique. Other more advanced approaches aim to produce digital content using exclusively SL. The SignLinking system (Fels, Richards, Hardman, & Lee, 2006), for example, introduces an interaction modality featuring hyperlinks embedded within a video. They are defined as SignLinks. Each SignLink spans a time window within the video. An icon typically indicates the presence of a link. As the icon appears, the user can choose whether to follow the link or to keep watching the video.

The above, and similar current approaches, to make digital content accessible to the deaf appear effective and technologically smart and attractive. However, they share the same drawbacks. Even if many deaf communities have recently replaced many functions of writing by using signed videos, this is not always either possible or appropriate. As an example, some typical actions on the Web are still not possible and technically difficult to implement by video: to take notes, or to annotate a Web resource (tagging), or to enter a query on a search engine. Actually, the first two mentioned actions could be (partially) supported by a multimodal annotation tool (see for example Bottoni, Levialdi, Pambuffetti, Panizzi, & Trinchese, 2006). Nevertheless, as searching is involved, and most of all searching according to an intended "meaning," we again encounter all the problems related to automatic interpretation and translation. Furthermore, videos are not anonymous: anyone can recognize the contributor simply looking at the video…unless the contributor wears a mask. This holds many people back who would otherwise be eager to contribute. Finally, people cannot easily edit or add further content to a video that someone else has produced, so a wikilike Web site in SL is not possible (Efthimiou et al., 2010). In summary, videos lack the ease of handling and the variegate usability of a written expression. Moreover, using signing avatars is not free from limitations. Automatic translation/transduction from text to a "convincing" signed sequence produced by an avatar is quite unfeasible, given the above discussion. Notwithstanding the many papers in the literature (see Wolfe et al., 2015), there is still a long way to go (Kipp, Nguyuen, Heloir, & Matthes, 2011), even from the point of view of user acceptance. Moreover, carrying out the conversion in advance raises the same limitations of video as it cannot be used for any normal real-time activity on the Web.

We can conclude that attempts for automatic interpretation as well as "translation" are at a very early and immature stage, and also using signed videos presents operational and accessibility limitations. As a consequence of these observations, the inclusion of written SL (and, as a consequence, the informatization of SL) rises to paramount importance in order to achieve an effective deaf-oriented accessibility design, and to ultimately mitigate the impact of the digital divide on deaf people (Borgia, Bianchini, Dalle, & De Marsico, 2012). We therefore embrace the project of supporting the ease of use of written forms of SL also in digital settings, to allow easy production and sharing of digital documents in a suitable notation. This would definitely improve accessibility for deaf users. In particular, our framework is based on the use of SignWriting, one of the proposed writing systems for SL. Some of these systems will be shortly presented in the next section.

Writing Systems for SL in Literature

The method traditionally used by linguists to deal with languages is glossing. According to the main intended use, a *gloss* is a brief marginal notation of the meaning of a word or wording in a text. It may be produced either in the language of the text, or in the language of the reader, if they are different. Therefore, *glossing* is also writing one language in another, where the written information is the *gloss*. In the case of SL, it entails looking at someone signing and writing sign by sign, by further including various notations to account for the facial and body grammar that goes with the signs. In practice, there is no attempt at interpretation but instead transcription is the

```
_____hu  ___________________________q  ______________
AWFUL, KNOW++ C-H-A-L-L-E-N-G-E-R, [5,R]-CL•surface
_______________pc                       ________________
pass under rocket•, SUDDEN-WRONG(conj) EXPLODEwg•slow
___pow
motion•. DISGUSTING.
```

Figure 33.2 Example of glossing for ASL.

objective. *Glosses* are labels in verbal language that are added to express the meaning of a SL expression in a very simplified way. Due to their popularity among linguists, they have long been a very common way of representing SLs, possibly adding annotations with further generally alphabetical symbols. The use of *glosses* is often rather considered a pseudonotation (Bianchini, 2012), while true notation forms have been attempted along time. This is also reflected in the considerations of Wilcox and Wilcox (1997) regarding American Sign Language (ASL), which also apply to any SL: "Glosses are rough translations of ASL morphemes into English morphemes. The reader should understand, however, that the existence of glosses for ASL does not signify that ASL is English. The reader should also remember that glosses are not intended to be good translations into English of these ASL sentences." Glossing SL means writing down a series of English words intended to represent signs (their "names") in the order they would be produced in ASL. Words are annotated by additional information about the way each sign is made. Such information includes direction, kind of motion, possible repetition, and nonmanual features, and is expressed by some standard symbols, or simply by describing the inflection used. An example is shown in Figure 33.2. A horizontal line placed over the gloss indicates simultaneous elements of ASL, while nonmanual features are described above the line.

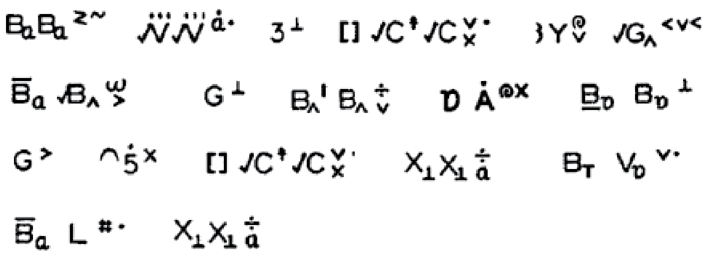
It should be understood that symbols used in glossing are rather related to the sentence construction, for example prosody, grammar, or repetition, while the information on the form of the sign is lost (see Antinoro Pizzuto, Chiari et al., 2010).

Devising a writing system especially dedicated to a language implies to understand its inner structure first. As reported above, one of the first researchers to appreciate the real linguistic and communication potential of SLs was William Stokoe (1960). His seminal work still provides plenty of insight on the issue, so that is has been reprinted (Stokoe, 2005). He gave a first fundamental contribution toward recognizing the richness and expressiveness of SLs, and one of the first notations used to "write" it. However, the limitations of such notation are twofold. First, it focuses only on the manual component of gestures. Stokoe's four-parameter model is characterized by symbols that represent hand shape, hand orientation, location relative to other parts of the body, and movement. Second, it is much more appropriate for use in linguistic studies than in every day communication. This can be easily verified by looking at Figure 33.3.

Last but not least, Stokoe notation basically relies on studies on ASL, and in this respect is not fully applicable without modifications for other SLs.

An attempt to preserve better the visual information contained in signs is the Hamburg Notation System (HamNoSys) (Hanke, 2004). It was created in 1985 from the initiative of a group of researchers at the University of Hamburg. It is based on Stokoe's four-parameter model. Each sign (or word) is written by assigning a value to each of these parameters. Its further revisions attempt to allow any signed language to

(a)



(b)

$B_aB_a^{z\sim}$	$\sqrt{\ddot{V}}\sqrt{\ddot{V}}\ \dot{a}\cdot$	$3^{\perp}$	$[]\sqrt{C}^{\ddagger}\sqrt{C}^{v\cdot}_{x}$	$3Y^{@}_{v}$	$\sqrt{G}_{\wedge}^{<v<}$
STORY(?)	QUOTE	THREE	BEAR(S)	GOLDILOCKS	WAY.IN

The story "Goldilocks and the Three Bears". Deep in

$\overline{B}_a\sqrt{B}_{\wedge\omega}$	$G^{\perp}$	$B_{\wedge}{}^{\prime}B_{\wedge v}^{\div}$	$D\mathring{A}^{@x}$	$\underline{B}_DB_D{}^{\perp}$
WOODS	UP	HOUSE	SITTING.THERE	ENTER

the woods, there is a house sitting on a hill. (If you) go in,

$G^{>}$	$\frown 5^{x}$	$[]\sqrt{C}^{\ddagger}\sqrt{C}^{v\cdot}_{x}$	$X_{\perp}X_{\perp a}^{\div}$	$B_TV_D^{v\cdot}$	$\overline{B}_aL^{\#\cdot}$	$X_{\perp}X_{\perp a}^{\div}$
THAT.THERE	FATHER	BEAR	OPEN.PAPER	READ	NEWSPAPER	OPEN.PAPER

(you will see) there Papa Bear reading the paper.

Figure 33.3 (a) A passage from Goldilocks in ASL transcribed in Stokoe notation (from Wikipedia). (b) Glossing of the same passage (from http://scriptsource.org/).

be written precisely, therefore overcoming some limitations of Stokoe notation. In addition, the writing system includes some support for nonmanual features. The writing symbols are available as a unicode-based character set for various operating systems, and can be downloaded for free. The language is also a basis for a series of avatar controls. However, even HamNoSys was not devised for writing full sentences. It therefore continues to be popular for academic purposes and has undergone four revisions but it is still unfeasible to use it as a script for everyday common communication. Figure 33.4 shows an example of the notation that also exemplifies the lack of strict visual correspondence with the sign. It is not important to give the meaning of the sign, which might be different across SLs, but rather we can consider the ability of the notation to immediately evoke the shape of the sign. In this regard, we can mention the distinction made by psychologists between two types of memory retrieval, according to the very well-known principle of recognizing versus recall. Recognition refers to our ability to recognize something familiar, while recall entails the retrieval of related information from memory. The prevalence of the former over the latter positively influences learnability.

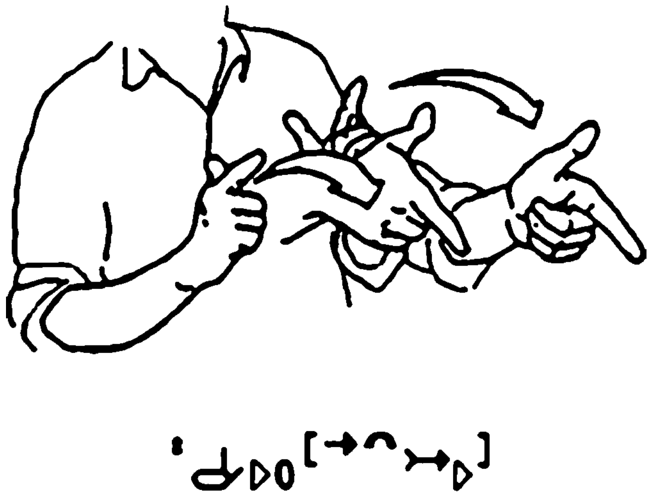


Figure 33.4 A sign and its transcription in HamNoSys notation.

While the basis for writing systems of VLs is the learned acoustic correspondence, the key for SL is to devise a visual correspondence, since the visual channel is exclusively used to communicate. Of course giving a full listing of possible notations for SL is out of the scope of this chapter. The interested reader can refer for example to the interesting web site https://aslfont.github.io/Symbol-Font-For-ASL/

The above considerations lead us to consider one of the notations that are attracting a very wide interest in both the deaf community and among linguists carrying out research on SL, namely SignWriting.

Of course, the choice of a writing system is subordinated to the goal that one must pursue. For the point of view of human-computer interaction, SignWriting proves an extremely appropriate candidate to work with because it has features that can rarely be found in modern SL writing systems. The system features a high level of iconicity, which in turn makes it very easy to learn and to master. Finally, the possibility of being employed in everyday use makes it the ideal candidate for a wide diffusion (Borgia et al., 2014).

SignWriting (Sutton, 1977, 1995) is an iconic writing system for SLs. In SignWriting, a combination of two-dimensional symbols, called glyphs, is used to represent any possible sign in any SL. The glyphs are abstract images depicting positions or movements of hands, face, and body. Figure 33.5 shows the LIS sign for "fun," written in SignWriting. Apart from the actual meaning of the sign, and from the use of a few conventions quite easy to grasp, the immediate evocation of the gestures used to produce the sign is instantly noticeable.

The high iconicity of this system is due to the shapes of the glyphs themselves, which have been conceived to reproduce any movement or position of the upper part of the human body, in a stylized yet accurate way. Sign languages are characterized by the three-dimensional spatial arrangement of gestures and by their temporal structure; in the very same way, the spatial arrangement of the glyphs in the page plays a core role. In fact, it does not follow a sequential order (like the letters of the written form of VLs), but the natural arrangement suggested by the human body.

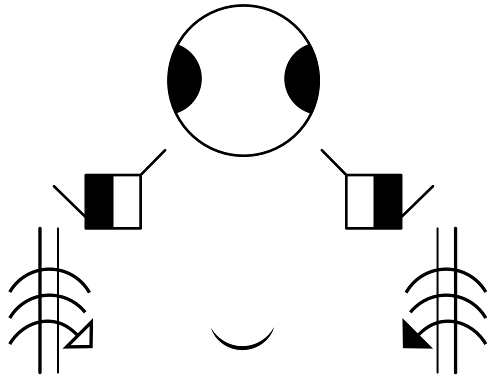


Figure 33.5 Italian sign language sign for "fun," written in SignWriting.

As SignWriting represents the actual physical formation of signs rather than their meaning, no phonemic or semantic analysis of a language is required to write it. A person who has learned the system can "feel out" an unfamiliar sign in the same way an English speaking person can "sound out" an unfamiliar word written in the Latin alphabet, without even needing to know what the sign means. Since 1996, the SignWriting standard also recommends writing following a vertical organization: writing signs in columns, one below the other. For this reason, most of the SignWriting texts available are written adopting this organization.

The set of movements and positions that a human body can produce from the waist up is huge. As a consequence, the set of glyphs that SignWriting provides to write down any sign is accordingly vast (about 30,000 units). The whole set of glyphs is referred to as the International SignWriting Alphabet (ISWA) (Slevinski, 2010a). The ISWA organizes the glyphs by dividing them into seven categories, identified by following a very intuitive principle: each one covers a different anatomic part of the human body, with a small number of exceptions. Further distinctions, i.e. groups and base symbols, are present within each category. This helps keeping a logical and linguistic organization within categories, which would otherwise be too vast to manage. Categories, groups, base symbols, and variations allow a unique code (ISWA code) for each glyph within the ISWA to be identified. Such a code is a key element for the digitalization of SignWriting because it is much easier for a machine to work with 13-digit codes, rather than with raw unorganized symbols.

Digital SignWriting

Aim and available software

A pencil and a piece of paper are the only required items to produce signs using SignWriting. However, the system has already risen as an effective means of communication for deaf people in the digital world, thanks to a 30-year informatization process started in 1986 by Richard Gleaves with the SignWriter computer program (Sutton, 1995).

First of all, there are bilingual websites or blogs (e.g. Frost, 2006) accessible from both hearing and deaf users by supporting both VL—in English—and American Sign

Language (ASL). Furthermore, unlike signed videos, the inclusion of written SL also enables the creation of wikilike websites. In fact, an ASL Wikipedia project (Sutton, 2012) is currently underway. The ASL articles in the ASL Wikipedia are written by deaf users, mainly as ASL translations of VL articles in the English Wikipedia. The goal of the project is to provide a bilingual, educational, informational tool intended to "provide information about the world to deaf and hearing people who use ASL as their daily, primary language" (Sutton, 2012).

Dictionaries and sign databases are among the online resources available in SignWriting. Such repositories as SignBank provide SL users with an archive of illustrations, SL videoclips and animations. Most importantly, SignBank provides a way to find words by looking up signs, using SignWriting symbols or VL keywords. This kind of digital artifact may prove a valuable asset for those who use Sign Language on a daily basis (especially for content authoring), whether they be deaf or hearing people.

Finally, most SignWriting digital resources are mainly available only thanks to a specific class of software, i.e. SignWriting digital editors. Such applications are the tools that enable the creation of digital signs written in SignWriting. In other words, they are critical for the informatization of SignWriting. Many applications have been produced by different research teams. Sutton's official SignMaker editor (Slevinski, 2010b) is one of the most popular, but a fair number of alternatives are available, such as SWift, DELEGs and SignWriter studio. Most SignWriting digital editors basically provide the same functionalities. Despite differences in interface design and implementation existing from one editor to another, such functionalities are:

- search for one or more glyphs belonging to the ISWA;
- insert the chosen glyph(s) onto an area designated for the composition of the sign;
- save the sign in one or more formats.

In general, the informatization of a writing system for SL poses different challenges. First of all, computer scientists need to devise effective and efficient ways of dealing with a set of symbols as large as the set of movements and positions that can be produced from the waist up. When designing a SignWriting digital editor, for instance, the large cardinality of the ISWA set might become a major problem for the application. If addressed incorrectly, it might affect both logic and presentation layers. It is necessary to get as close as possible to the "aurea mediocritas" between the unrestricted access to the data (the glyphs) and the presentation of a human-manageable amount of information. Furthermore, the rules underlying the composition and the organization of a SL writing system are (generally) very different from those of a VL system. For instance, SignWriting grants a very high degree of freedom to its users. There is no high-level rule at all about the composition itself: no restriction is set to the number of glyphs within a sign, to their possible spatial arrangement and to their relative positioning.

A digital editor for SignWriting: SWift

In this section, we introduce SWift (SignWriting improved fast transcriber) (Bianchini, Borgia, & De Marsico, 2012), an editor conceived by a research team of (both deaf and hearing) linguists, SignWriting users, and computer scientists in the field of HCI. SignWriting editors are seldom designed with usability as the main focus. In most cases the main goals are giving the user an unrestricted access to the glyphs, and providing the necessary functionalities to manage the created sign. On the other hand, SWift has been designed in close collaboration with a true sample of its target users, namely the hearing and deaf researchers of CNR ISTC (Istituto di Scienze e Tecnologie della Cognizione—Institute for Science and Technologies of Cognition), in full compliance with the core principles of user-centered design (Norman & Draper, 1986) and contextual design (Wixon, Holtzblatt, & Knox, 1990). Both design techniques require the involvement of final users as the main stakeholders in the design process. The first technique requires giving extensive attention to needs, preferences, and limitations of the end users of a product. This must be done at each stage of the design process. The second technique requires that researchers aggregate data from customers in the very context where they will use the product, and that apply their findings during the design stages. Actually, a number of further stakeholders should be involved when designing for people with special needs (De Marsico, Kimani, Mirabella, Norman, & Catarci, 2006). These include: experts in specific disabilities, which can suggest the best ergonomic strategies to support these users; experts in the target application domain, which can suggest alternative channels and/or message modifications to better reach these users; and experts in the accessibility guidelines regarding users with special needs. It could prove very helpful to follow some specific requirements, while designing the User Interface (UI) of an accessible application for deaf people, for example a digital editor like Swift. They are listed below:

- Intuitiveness: the user should be relieved from the burden of learning the UI, whereas the UI can simply be understood; with this purpose in mind, each function should be presented (and work) in a intuitive and familiar way.
- Minimization of information: each screen should present a small amount of essential information, in order to avoid overwhelming the user with a cluttered UI.
- Look and feel: icons should be simple, large, and familiar; if their meaning remains unclear, mouse-over-triggered animations/videos could be embedded in the buttons/links to guide the user; dealing with text labels in the UI might be difficult for deaf people, therefore such elements should be kept at a minimum (Perfetti & Sandak, 2000).
- User-driven interface testing: each change (or the most important changes at least) in the UI should be discussed and tested with a team including the target users of the application; the testing should involve high-level aspects as well as low-level ones (such as the spatial placement of buttons within the UI).

The UI of SWift (shown in Figure 33.6) is an example of deaf-accessible UI, which has been designed meeting all of the requirements mentioned above. Like other digital editors, SWift provides an area to compose the sign. Such a component is referred to as the *sign display*; it is a whiteboard-resembling area whose purpose is to show the sign that the user is currently composing. A sign is composed by dragging a number of *glyphs* from the *glyph menu* and dropping them on the *sign display*. Once they are placed there, they become both draggable (to be relocated at will) and selectable (for editing purposes). The far larger size of the *glyph menu* with respect to the *sign display* may seem strange but the reason for this will be clear soon.

The *glyph menu* allows the user to search any glyph within the ISWA. Once the user finds the desired glyph, he/she can drag it and drop it on the *sign display* in order to

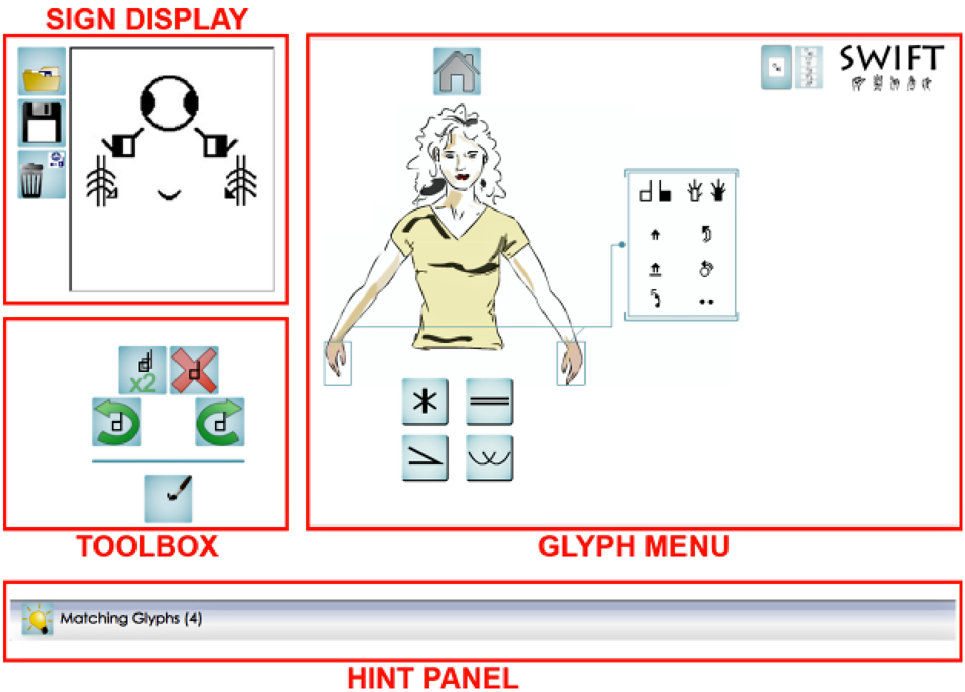


Figure 33.6 Home screen of SWift, divided into four functional areas.

include it within the sign that is under construction. Most efforts were devoted to make the interaction with the *glyph menu* fast and effective. The underlying concept is basically "Why browse, when you can search?" For this reason, the *glyph menu* features a search system that implements a completely different approach with respect to its competitors (Bianchini, Borgia, Bottoni, & De Marsica, 2012). In order to support SignWriting beginners during the composition, the *glyph menu* has been designed to present a stylized human figure (Figure 33.6), folksily called *puppet*, as a starting point for the search of a glyph. The purpose of the *puppet* is making the search of a glyph easier and faster, by making the choice of anatomic areas (categories and groups) more straightforward. In other words, it enforces recognition versus recall. The buttons below the *puppet* represent groups of glyphs related to signing aspects apart from body parts, for example, repetitions and contact points. By choosing an anatomic area of the *puppet*, or one of the buttons below, the user accesses the dedicated search menu for that area or for that kind of item. After the user clicks, the *puppet* and the buttons beneath are reduced and shifted to the left, but remain clickable and form a navigation menu together with the button to return to the *glyph menu*'s home screen. This allows freely navigating from one area to another. A red square around the selected area reminds the user's choice, like breadcrumbs. In the central part of the menu, a label and an icon show what kind of glyphs are available using the group of boxes beneath. These boxes are referred to as *choice boxes* and guide the user during the search for a glyph. They display groups of options: the user can choose only one element from each of them, in any order, and each choice progressively restricts the set of symbols displayed as response to the user search. Figure 33.7 shows an

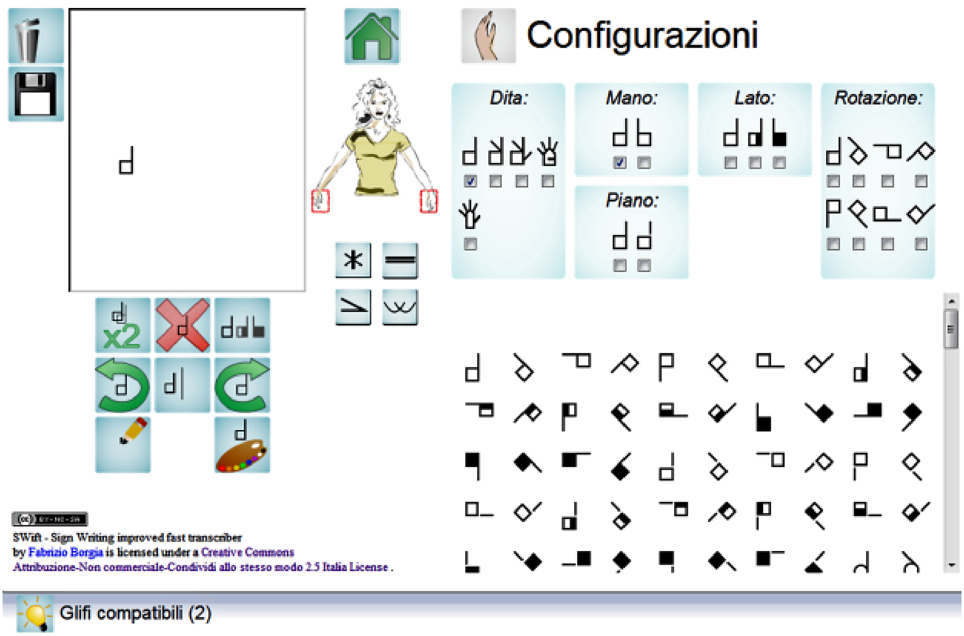


Figure 33.7 An example of navigation support.

example. The configuration of the *glyph menu* at any step explains the need for a large area in the interface. Once a glyph is found, it can be dragged and dropped directly on the *sign display*.

Finally, the *hint panel* is one of the innovations that distinguishes SWift from other digital editors because it implements a prototype predictive sign composition for SL. Many studies, such as Curran, Woods, and O'Riordan (2006), demonstrate that predictive text is an important aid to communication when handling the set of characters of a VL, which are in the order of magnitude of tens. It is easy to realize how this can improve the interaction with a set of tens of thousands of symbols like ISWA. The *hint panel* enables predictive sign by showing in real time, as the composition process is underway, a set of glyphs that are *compatible* with those the user already entered in the *sign display*. Compatibility is computed and updated according to the rate of cooccurrence of the glyphs in a database of known and already stored signs. The glyphs suggested in the *hint panel* are immediately available to be inserted into the *sign display*. With such action, the user can save the effort and the time required to search for each glyph from scratch.

Assessing usability of applications for deaf users

Assessing the reliability and the usability of a deaf-accessible application is of paramount importance during the development lifecycle. It is advisable, yet not always possible, to conduct test sessions with participants representing the actual end users of the application. In the case of a deaf-accessible application, deaf people should be included within the group of the participants; in the particular case of a digital editor (such as SWift) it was necessary to test the application with deaf people being also proficient in SignWriting. However, whenever very specific requirements have to be met to reflect the real target community, a statistical sample in the order of tens is often the best that one can achieve. For instance, this was the case with SWift.

The choice of the proper set of tools is very important during the usability assessment of an application. Most tools, however, do not consider the possibility that the sample is composed of people with special needs, such as deaf people. Currently, a very restricted set of tools and studies is available. One of the most popular and effective tools for usability testing is the think-aloud protocol (TAP) (Lewis & Rieman, 1993). However, since deaf participants cannot actually "Think Aloud," different research teams around the world adapted TAP in order to include SL. The first adaptation was suggested by Roberts and Fels (2006), who performed a number of tests with deaf users adopting a TAP-based protocol. Along the same line, an adaptation of the TAP, called Think by Signs (Bianchini, Borgia, Bottoni et al., 2012) is described here. The Think by Signs is a bilingual (VL and SL) test methodology, so it can be employed with both hearing and deaf people. The TAP itself partly interrupts the attention flow of the user because it engages cognitive resources along a different track. However, the positive aspect is the possibility of expressing one's own impressions in real time, without the possible bias due to the final outcome. In the specific case of deaf people, the interruption of the flow of the attention is even more concrete: the participant will inevitably stop using the keyboard and / or mouse because one or two hands are required in order to produce signs. However, signing during different actions is typical of the way deaf people have to communicate while performing a task, so this does not affect the outcome of the test.

More specifically, the Think by Signs test consists of two moments. The *welcome time* starts as soon as the participant sits down in front of the computer. The system displays a welcome screen containing a signed video (on the left part of the screen) and its VL translation (on the right part). Consistent with the rules of the TAP, the greeting sequence contains: a brief thanks for the participation, an explanation of the structure and the rules of the test, and finally a reminder about the purpose of the test (which is not conceived to test the skills of the participant but rather to test the capabilities and the usability of the software). The latter should help the participant in feeling at ease with the test. The *test time* follows the welcome time, and it represents the core part of the procedure. The participant is required to perform a list of tasks to test the functions of SWift and their usability. As stated by the rules of the TAP, during this phase the participant is asked to sign anything that comes to his / her mind. Given the possible high number of tasks, it is appropriate to alternate, when possible, simple assignments with complex ones, to avoid tiring the participant. For the same reason, the participant might need a reminder about the task that is currently under way. To address such a need it is recommended to design a task list to guide the participant. Several options are available to create a bilingual task list. In the first place, the test designer should decide whether to delegate the SL inclusion to an electronic device (with signed videos illustrating each task) or to an interpreter. In the second place, consistent with the first choice, the specific role and responsibilities of the device / interpreter must be clearly defined. In this regard, Borgia (2015) observed that the involvement of an interpreter usually makes the user feel more comfortable than a recorded explanation. This is also due to the possibility of directly or indirectly asking questions about the required activities, and, of course, it increases the probability of a correct understanding of the tasks too. For this reason it is advisable to choose an option involving an interpreter. Given this, it is further necessary to define the way the interpreter has to act during the test. One of the best options is to have the interpreter always provide an initial SL translation of the task. Therefore, at the beginning of each task, a card is presented to the participant. The card contains a simple, direct VL question that identifies the task. As the card is presented, the interpreter signs the question to the participant. Afterwards, the participant is allowed to ask questions to the investigator, using the interpreter as intermediary. The answer may follow only if it does not affect the outcome of the test. Last but not least, deaf people disclose their attitudes and sensations through a very high variability of SL nonmanual components. Due to this characteristic, it is of paramount importance to capture the user's face when recording a test session. Different spatial settings for the test room are available in Roberts & Fels (2006) and Borgia (2015).

In compliance with the guidelines discussed above, the sample of participants to the usability assessment of SWift was mainly composed by deaf people, and a Think by Signs test was carried out. At the end of the test, the participants were also asked to fill in a questionnaire to assess their subjective satisfaction with specific aspects of the UI. The tool chosen for the job was the Questionnaire for User Interaction Satisfaction (QUIS) (Chin, Diehl, & Norman, 1988). Like the TAP, even the QUIS was adapted to fit in a bilingual test environment, by presenting any suitable question among the set of original ones (and any possible answer) both in VL and SL with the aid of signed videos.

Optical glyph recognition for SignWriting: SW-OGR

Despite the efforts of different research teams, SignWriting digital editors are still far from granting the user an interface that is able to emulate the simplicity of handwriting. Actually, any software solution developed to support the production of SignWriting documents relies heavily on Windows, Icons, Menus, Pointer (WIMP) interfaces, both for accessing the application features and for the sign composition process itself. The problem is all but a theoretical one. Even dealing with word processors (Latin alphabet), the users often feel the higher complexity and the slower composition time with respect to handwritten text production. Given its huge number of glyphs, this especially holds for people using SignWriting, in particular for deaf people. In fact, they are far more accurate, fast, and comfortable using the paper-pencil approach rather than dealing with the (more-or-less) complex interaction styles of a digital editor. For this reason, the design of a new generation of SignWriting editors has been planned (Borgia et al., 2014), able to relieve the user of any, or most at least, burden related to clicking, dragging, searching, browsing on the UI during the composition process of a sign. The purpose of the designers is to implement an interaction style that is as similar as possible to the paper-pencil approach that humans normally use when writing or drawing. In order to achieve such a goal, it is necessary to integrate the digital editor with another software module (see Figure 33.8)—more specifically an Optical Character Recognition (OCR) engine. The purpose of this additional module is to operate the electronic conversion of images, containing handwritten or printed SignWriting symbols, into ISWA-encoded SignWriting texts. Such technique is known as the SignWriting Optical Glyph Recognition (SW-OGR) (Borgia, 2015).

Of course, since WIMP is currently the easiest, most common interface style in the world, it cannot be totally left behind, because it is necessary to access the

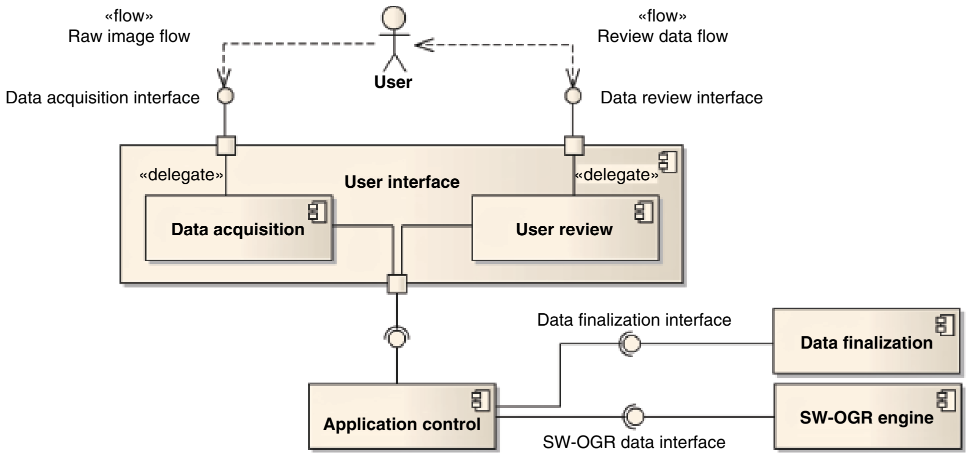


Figure 33.8 Component diagram for a new generation of SignWriting editors featuring a SW-OGR engine.

features of most applications. Nevertheless, dismissing the WIMP style during the sign composition, which is the core part of any SignWriting editor, could prove a rewarding choice.

A conceptual schema of the new generation of OGR-powered SignWriting digital editors is illustrated by the diagram in Figure 33.8; the application is composed by:

- User interface, which includes two subcomponents.
  - Data acquisition, which provides the user with a simple interaction style to compose signs, focusing on intuitiveness (or, better, transparency) and accuracy; this component is also designed to collect the data produced by the user (typically an image) and to pass it to the control component.
  - User review, which comes into play after the recognition; it allows the user to make corrections (no recognition is 100% accurate) and/or add further data.
- Application control, which implements the (model-view-controller pattern) controller of the application; among other tasks, this component also coordinates the data flow between the UI and the SW-OGR engine.
- SW-OGR engine, which is the model component of the application; its purpose is to provide a fast and accurate recognition of all (or most) glyphs handwritten or submitted by the user; the final product of the recognition are an image, and an associated data file containing the ISWA codes (and the coordinates) of the recognized glyphs.
- Data finalization module, which saves the user-reviewed data in the requested format (image file, XML file, database entries, etc.).

The SW-OGR engine performs the recognition of SignWriting texts by only working with the geometric and topological features of the symbols, and with their topological relationships. The recognition also takes advantage of context-dependent information, such as the knowledge of the structure of the ISWA—its categories, groups, and so forth. The engine is intended to serve a twofold purpose: first, it can be embedded within existing SignWriting editors, such as SWift, in order to provide prompt support for handwriting, and make the composition process much faster and more comfortable for everyday use. In addition, it is worth noting that a considerable number of paper-handwritten SignWriting corpora exist, gathered from different communities around the world. Those corpora are an invaluable asset, and they could become even more useful if digitalized.

Conclusions and Integration Perspectives

In the context of deaf-oriented accessibility, applications to support written SL, whether they are digital editors or highly specialized applications (like the SW-OGR engine) are not intended to be separated realities. In the specific case of SignWriting, the integration perspectives of SignWriting digital editors and SW-OGR are very interesting. In fact, observing the diagram in Figure 33.8, it is possible to infer that the application handling the UI (and the application control) could be one of the already existing digital editors, because the required features and application interfaces substantially correspond. Since any application concerning digital SignWriting shares the same way of representing signs, the integration is easy to implement. In fact, both digital editors and SW-OGR represent a sign as an XML document using the SignWriting Markup Language dialect (SWML). Within the document, each sign is associated to the list of its component glyphs, storing their ISWA codes and their spatial coordinates in the sign space (like the Sign Display in Figure 33.6). As a consequence, a document produced by SW-OGR engine can be read and updated using SWift. Such interoperability is the idea underlying a multiapplication framework, whose purpose is making SignWriting effectively exploitable as a communication mean and as a learning support for deaf people. Such framework, named SignWriting-oriented resources for the deaf (SWord), has already been sketched (Borgia et al., 2014), and is under development. SWord is intended to support the acquisition of a corpus of signs from two possible sources: user-composed via digital editors (SWift, in particular), and digitized SignWriting corpora (currently on paper) via SW-OGR. An intermediate goal of the framework is to use these acquisition methods to gather a significant amount of signs, to be stored in electronic form together with their decomposition into glyphs. A set of signs of this kind is referred to as a *structured corpus* (Borgia, 2015). The purpose of a corpus prepared in this way is to allow the identification of recurring patterns in the composition of the signs, and the computation of relevant statistics on the transcribed form of the signs. These elements are of paramount importance in order to gain a deeper understanding of the rules of SLs. Ultimately, the precise linguistic and production information stored by each SignWriting glyph can allow computer scientists to use them to determine the movements and expression of a signing avatar in a very accurate and satisfactory way. A similar approach has been already adopted by Karpov and Ronzhin (2014) by implementing a 3D signing avatar for Russian sign language, which is based on signs represented with HamNoSys. In this way, one might avoid using written text, which is almost impossible to translate automatically in SL, to derive the avatar behavior. Using the intermediate form of SignWriting as an alternative starting point to guide the avatar, people using Sign Language may be supported in a number of activities that would otherwise require the use of a VL (e.g. e-learning), even without directly knowing SignWriting.

Finally, "transduction" of gestures from signed videos into SignWriting documents is the final step in the plan for the overall SWord. This step is at present at a very early stage, due to difficulties underlying computer vision-based approaches to the problem. These difficulties are mostly due to occlusion and self-occlusion of relevant body parts during signing, and to the "tokenization" of a visual sequence. However, we are strongly convinced that it is worth devoting more efforts to this and similar projects, to provide full expressive possibilities in the digital world even to deaf people.

In Memory

We would like to remember Elena Antinoro Pizzuto and Patrice Dalle, who, with their great enthusiasm, enormous competence, and scientific sensibility, spurred research as well as debate on the deaf world and on its fascinating language.

Acknowledgements

We thank the group of hearing and deaf researchers of CNR ISTC (Istituto di Scienze e Tecnologie della Cognizione—Institute for Science and Technologies of Cognition), because working with them led us to a deeper comprehension of deaf problems and sensibility. Thanks to their collaboration, our framework is taking shape and substance.

References

Antinoro Pizzuto, E. (2008). Meccanismi di coesione testuale e strutture di grande iconicità nella lingua dei segni italiana (LIS) e altre lingue dei segni. In *Atti del convegno nazionale "La grammatica della Lingua Italiana dei Segni"* (pp. 137–158). Venice, Italy: Cafoscarina.

Antinoro Pizzuto, E., Bianchini, C. S., Capuano, D., Gianfreda, G., & Rossini, P. (2010). Language resources and visual communication in a deaf-centered multimodal e-learning environment: issues to be addressed. In *Workshop on Supporting eLearning with Language Resources and Semantic Data. Proceedings of Language Resources and Evaluation Conference 2010, Seventh International Conference on Language Resources and Evaluation* (p. 18–23). Paris: European Language Resources Association.

Antinoro Pizzuto, E., Chiari, I., & Rossini, P. (2010). *Representing signed languages: Ttheoretical, methodological and practical issues.* In M. Pettorino, F. A. Giannini, I. Chiari & F. Dovetto (Eds.), *Spoken communication* (pp. 205–241). Newcastle upon Tyne: Cambridge Scholars Publishing.

Antinoro Pizzuto, E., Rossini P, Sallandre, M.-A., & Wilkinson, E. (2008). Deixis, anaphora and highly iconic structures: Cross-linguistic evidence on American (ASL), French (LSF) and Italian (LIS) signed languages. In M. de Quadros (Ed.) *Sign languages: Spinning and unraveling the past, present and future* (pp. 475–495). Petrópolis, Brazil: Editora Arara Azul.

Bianchini, C. S. (2012). *Analyse métalinguistique de l'émergence d'un système d'écriture des Langues des Signes: SignWriting et son application à la Langue des Signes Italienne (LIS)* (Unpublished PhD dissertation). University Paris VIII, Paris.

Bianchini, C. S., Borgia, F., Bottoni, P., & De Marsico, M. (2012). SWift: a SignWriting improved fast transcriber. In G. Tortora, S. Levialdi & M. Tucci (Eds.), *Proceedings of the International Working Conference on Advanced Visual Interfaces* (pp. 390–393). New York, NY: ACM. doi:10.1145/2254556.2254631

Bianchini, C. S., Borgia, F., & De Marsico, M. (2012). SWift A SignWriting Editor to Bridge between Deaf World and E-learning. In *2013 IEEE 13th International Conference on Advanced Learning Technologies* (pp. 526–530). Los Alamitos, CA: IEEE Computer Society.

Borgia, F. (2015). *Informatisation d'une forme graphique des Langues des Signes: application au système d'écriture SignWriting* (Unpublished PhD dissertation). Universite Toulouse III—Paul Sabatier, Toulouse, France.

Borgia, F., Bianchini, C. S., Dalle, P., & De Marsico, M. (2012). Resource production of written forms of Sign Languages by a user-centered editor, SWift (SignWriting improved fast transcriber). In *International Conference on Language Resources and Evaluation—Language Resources and Evaluation Conference* (pp. 3779–3784). doi:10.13140/RG.2.1.5127.6646

Borgia, F., Bianchini, C., & De Marsico, M. (2014). Towards improving the e-learning experience for deaf students: e-LUX. In C. Stephanidis & M. Antona (Eds.), *Lecture Notes in Computer Science: Universal Access in Human-Computer Interaction* (vol. 8514, p. 221–232). Berlin: Springer. doi: 10.1007/978-3-319-07440-5 21

Bottoni, P., Borgia, F., Buccarella, D., Capuano, D., De Marsico, M., & Labella, A. (2013). Stories and signs in an e-learning environment for deaf people. *Universal access in the information society*, *12*(4), 369–386.

Bottoni, P., Levialdi, S., Pambuffetti, N., Panizzi, E., & Trinchese, R. (2006). Storing and retrieving multimedia web notes. *International Journal of Computational Science and Engineering*, *2*(5–6), 341–358.

Braffort, A., & Dalle, P. (2008). Sign language applications: preliminary modeling. *Universal Access in the Information Society*, *6*(4), 393–404.

Caldwell, B., Cooper, M., Guarino Reid, L., & Vanderheiden, G. (2008). *Web content accessibility guidelines (WCAG) 2.0.* World Wide Web Consortium. Retrieved from http://www.w3.org/TR/WCAG20/

Chi, E. H., Pirolli, P., Chen, K., & Pitkow, J. (2001). Using information scent to model user information needs and actions and the Web. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 490–497). New York, NY: ACM.

Chin, J. P., Diehl, V. A., & Norman, K. L. (1988). Development of an instrument measuring user satisfaction of the human-computer interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 213–218). New York, NY: ACM.

Curran, K., Woods, D., & O'Riordan, B. (2006). Investigating text input methods for mobile phones. *Telematics and Informatics*, *23*(1), 1–21.

Cuxac, C. (1996). *Fonctions et structures de l'iconicité des langues des signes* (Unpublished thesis). Paris: Université Paris V.

Cuxac, C. (2000). *La Langue des Signes Française (LSF): les voies de l'iconicité.* Paris: Ophrys.

Cuxac, C. & Antinoro Pizzuto, E. (2010). Émergence, norme et variation dans les langues des signes: Vers une redéfinition notionnelle. In B. Garcia & M. Derycke (Eds.), Sourds et langues des signes: norme et variations [Special issue]. *Langage et Société*, *131*, 37–53.

Cuxac, C., Braffort, A., Choisier, A., Collet, C., Dalle, P., Fusselier, I.,…Sallandre, M.-A. (2002). *Corpus LS-COLIN.* Retrieved from http://cocoon.huma-num.fr/exist/crdo/meta/crdo-FSL-CUC021_SOUND

Cuxac, C. & Sallandre, M-A. (2007). Iconicity and arbitrariness in French Sign Language: Highly iconic structures, degenerated iconicity and diagrammatic iconicity. In E. Antinoro Pizzuto, P. Pietrandrea & R. Simone (Eds.), *Verbal and signed languages: Comparing structures, constructs and methodologies* (pp. 13–33). Berlin: Mouton de Gruyter.

De Marsico, M., Kimani, S., Mirabella, V., Norman, K. L., & Catarci, T. (2006). A proposal toward the development of accessible e-learning content by human involvement. *Universal Access in the Information Society*, *5*(2), 150–169.

Efthimiou, E., Fotinea, E., & Glauert, J. (Eds.). (2008). Emerging technologies for deaf accessibility in the information society [Special issue]. *Universal Access in the Information Society*, *6*(4).

Efthimiou, E., Fotinea, S., Hanke, T., Glauert, J., Bowden, R., Braffort, A.,…Goudenove, F. (2010). DICTA-SIGN: Sign language recognition, generation and modelling with application in deaf communication. In *Proceedings of Language Resources and Evaluation Conference 2010, Seventh International Conference on Language Resources and Evaluation* (p. 80–83). Paris: European Language Resources Association.

Elliott, R., Glauert, J. R., Kennaway, J. R., Marshall, I., & Safar, E. (2008). Linguistic modelling and language-processing technologies for Avatar-based sign language presentation. *Universal Access in the Information Society*, *6*(4), 375–391.

Fajardo, I., Arfe, B., Benedetti, P. & Altoé, G. M. (2008). Hyper- link format, categorization abilities and memory span as contributors to deaf users hypertext access. *Journal of Deaf Studies and Deaf Education*, *13*(2), 241–256.

Fajardo, I., Parra, E., Cañas, J., Abascal, J., & Lopez, J. (2008, July). Web information search in Sign Language. In *Informal learning on the Web: Individual differences and evaluation processes.* Symposium conducted at the XXIX International Congress of Psychology, Berlin, Germany.

Fajardo, I., Vigo, M., Salmerón, L. (2009). Technology for supporting web information search and learning in Sign Language. *Interacting with Computers*, *21*(4), 243–256.

Fels, D., Richards, J., Hardman, J., & Lee, D. (2006). Sign language Web pages. *American Annals of the Deaf*, *151*(4), 423–433.

Frost, A. (2006). The Frost Village. Retrieved from http://www.frostvillage.com/

Garcia, B. & Sallandre, M.-A. (2014). Reference resolution in French Sign Language. In P. Cabredo Hofherr & A. Zribi-Hertz (Eds.), *Crosslinguistic studies on noun phrase structure and reference* (pp. 316–364). Leiden: Brill.

Gillot, D. (1998). *Le droit des sourds: 115 propositions. Report from House Representative D. Gillot to French Prime Minister L. Jospin.* Retrieved from http://maitre-autrive.com/contents/fr/rapport_gillot.pdf

Goldin-Meadow, S. (2001). How do profoundly deaf children learn to read? In *Learning, Disabilities Research and Practice*, *16*(4), 222–229.

Hanke, T. (2004). HamNoSys—representing sign language data in language resources and language processing contexts. In O. Streiter & C. Vettori (Eds.), *Language Resources and Evaluation Conference 2004, Workshop proceedings: Representation and processing of sign languages* (pp. 1–6). Paris: ELRA.

Karpov, A., & Ronzhin, A. (2014). A universal assistive technology with multimodal input and multimedia output interfaces. In C. Stephanidis and M. Antona M. (Eds.), *Universal access in human-computer interaction. Design and development methods for universal access.* Cham: Springer.

Kipp, M., Nguyen, Q., Heloir, A., & Matthes, S. (2011). Assessing the deaf user perspective on sign language avatars. In *Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility* (pp. 107–114). New York, NY: ACM.

Lewis, C., & Rieman, J. (1993). *Task-centered user interface design: A practical introduction.* Boulder, CO: University of Colorado.

Lewis, M. P., Simons, G. F. & Fennig, C. D. (Eds.). (2015). *Ethnologue: languages of the world* (18th ed.). Dallas, TX: SIL International. Retrieved from http://www.ethnologue.com/

Norman, D., & Draper, S. (1986). *User centered system design.* Mahwah, NJ: Lawrence Erlbaum.

Ong, S. C., & Ranganath, S. (2005). Automatic sign language analysis: A survey and the future beyond lexical meaning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(6), 873–891.

Perfetti, C., & Sandak, R. (2000). Reading optimally builds on spoken language. *Journal of Deaf Studies and Deaf Education*, *5*, 32–50.

Prensky, M. (2001). Digital natives, digital immigrants part 1. *On the Horizon*, *9*(5), 1–6.

Prillwitz, S., Leven, R., Zienert, H,. Hanke, T., & Henning, J. (1989). *Hamburg Notation System for Sign Languages: An introductory guide, HamNoSys version 2.0.* Hamburg: Signum Press.

Roberts, V., & Fels, D. (2006). Methods for inclusion: Employing think aloud protocols in software usability studies with individuals who are deaf. *International Journal of Human Computer Studies*, *64*(6), 489–501.

Salomoni, P., Mirri, S., Ferretti, S., & Roccetti, M. (2007). Profiling learners with special needs for custom e-learning experiences, a closed case? In *Proceedings of the 2007 International Cross-Disciplinary Conference on Web Accessibility (W4A)* (pp. 84–92). New York, NY: ACM.

Slevinski, S. E., Jr. (2010a). *International SignWriting Alphabet 2010 HTML reference.* Retrieved from http://www.signbank.org/iswa

Slevinski, S. E., Jr. (2010b). SignMaker. Retrieved from http://www.signbank.org/signpuddle2.0/signmaker.php?ui=1&sgn=63

Stokoe, W. C. (1960). *Sign Language structure: An outline of the visual communication systems of the American deaf.* Buffalo, NY, Department of Anthropology and Linguistics, University of Buffalo.

Stokoe, W. C. (2005). Sign language structure: An outline of the visual communication systems of the American deaf. Journal of Deaf Studies and Deaf Education, *10*(1), 3–37.

Sutton, V. (1995). *Lessons in SignWriting: textbook, and workbook.* La Jolla, CA: Deaf Action Committee for Sign Writing.

Sutton, V. (1977). Sutton Movement Shorthand: Writing tool for research. In W. C. Stokoe (Ed.), *Proceedings of the First National Symposium on Sign Language Research and Teaching* (pp. 267–296). Chicago, IL: Department of Health, Education and Welfare.

Sutton, V. (2012). *SignWriting encyclopedia projects.* Retrieved from http://www.signwriting.org/encyclopedia/

Wilcox, S., & Wilcox, P. P. (1997). *Learning to see: Teaching American Sign Language as a second language.* Washington, DC: Gallaudet University Press.

Wixon, D., Holtzblatt, K., & Knox, S. (1990). Contextual design: an emergent view of system design. In J. Carrasco & W. J. (Eds.), *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Empowering People* (pp. 329–336). New York, NY: ACM.

Wolfe, R., Efthimiou, E., Glauert, J., Hanke, T., McDonald, J., & Schnepp, J. (2015). *Universal Access in the Information Society* [Special issue: recent advances in sign language translation and avatar technology], *1–2.*

Part IX
Social

34

Social Networking

Jennifer Golbeck

Introduction

Online social networks have changed the way people interact with the Internet and with one another. They have dramatically reduced the effort required to maintain relationships that would have otherwise faded away, and they provide windows onto human behavior and relationships that have never been known in the past. Facebook alone, at the time or writing, has about 2 billion active monthly users. Instagram, Twitter, Pinterest, and Google + average around the 400 million mark. More than half of the world's population with Internet access is using some type of social networking. Studying online social networking builds upon the firm foundation of research into online communities that addressed users' interactions in the earlier days of the Internet and that carries on today (Preece, 2000).

At its heart, social networking is technology that lets people connect to others through an online medium. It typically has a mechanism where people explicitly create social relationships by friending, following, or the like (Golbeck, 2005). In the early days of online social networking, making those connections was often the core purpose but modern networks generally connect people to the posts and updates made by their social connections. Online social networks can also be built from implicit interactions, such as posts and replies in discussion groups or mentions in other forums.

This social interaction takes place through a computer-based technological medium, so much of the study of online social networking falls into the space of human-computer interaction (HCI). As a result, there is a broad and diverse set of topics that are important to understand. In this chapter, we break them down into broad categories. Social networking provides data that lets researchers understand human interaction and the world more broadly through the lens of what is shared online; the use of social networking sites is a phenomenon to study itself. There are design issues for the social networking platforms and visualizations to help us analyze the data within them. What people do allows researchers to build systems that improve interaction on a variety of platforms.

The Wiley Handbook of Human Computer Interaction, Volume 2, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

Understanding Humans

Access to extensive data on literally billions of people is unprecedented in human history. As a result, social media as a technology has been field changing when it comes to understanding people and how they behave and interact with one another. While social scientists and psychologists are increasing their work in this area, many human-computer interaction researchers have taken the first steps toward combining social science insights with social networking data to demonstrate how our understanding of humans can improve.

One of the key practices in this area of HCI is to draw on insights and theories developed in the social sciences before the advent of online social networking. Concepts like homophily (i.e. people tend to be friends with people similar to themselves), social capital, tie strength, and trust have all been mapped into a context that can be used in the online context. This follows in a long tradition within HCI of bringing existing theory into the HCI space, back to the early days of mapping Fitts' law to the graphical user interface (MacKenzie, 1992). Human-computer interaction research in this area can generally be broken down into two main categories: research focused on relationships and interaction, and research focused on individuals themselves.

Interactions and Relationships

Social capital is a concept that deals with the benefits produced from social relationships. It includes things like favors, trust, and support, and people "earn" this capital by helping others or the common good. It can be thought of as a type of power that people have to request or receive social support from others by earning their trust and support through other good acts. Ultimately, it is embedded in the social relationships between people.

This is one of the earliest concepts transitioned to and studied in online social networking. Early work compared the way that teens and adults used the early social networking site MySpace (Pfeil, Arjan, & Zaphiris, 2009), finding that teens had more friends but adults had more powerful friends. This was extended to a study of college students' use of Facebook and the importance of those relationships for social capital (Ellison, Steinfield, & Lampe, 2007). Work has even shown that online social network-based relationships can build social capital, especially among young people with low self-esteem who may otherwise have a difficult time establishing the important relationships (Steinfield, Ellison, & Lampe, 2008).

Tie strength is another sociological concept that has received attention in the social networking space. Popularized by Granovetter (1973), tie strength describes how close two people are, often measured as a combination of intimacy, favors done for one another, the time spent together, and emotional intensity. Gilbert and Karahalios (2009) demonstrated that the tie strength between two people can be predicted by analyzing their profiles and interactions on Facebook.

Trust relationships in online social networks have received quite a lot of attention by HCI researchers. Models for inferring trust between two individuals who are connected (Golbeck, 2009), and especially between those who are *not* directly connected (e.g. Avesani, Massa, & Tiella, 2005; Golbeck, 2005; Ziegler & Lausen, 2005) have been widely addressed. This research bridges HCI with artificial intelligence through the creation of algorithms that use this human data, understood through user studies and analysis, to build algorithms that automatically detect signs of trust.

Individual Attributes

Before using any data that people share online on social networking sites, there is a question about whether they share data that is an honest reflection of themselves. After all, if people are lying, there may not be much meaningful insight to glean from these networks. However, research has shown that users essentially present their true selves in their profiles (Back et al., 2010). With that in mind, there is a lot that can be done with the data people choose to share.

Online social networks have become a place to understand individual users more deeply by performing analysis that goes beyond what they have explicitly shared. Using data from social media profiles of thousands of users, researchers build models that allow for the inference of an attribute for any given individual. Traits are diverse and include sexual orientation (Jernigan & Mistree, 2009; Kosinski, Stillwell, & Graepel, 2013), race and religion (Kosinski et al., 2013), political preferences (Golbeck & Hansen, 2014), personality traits (Golbeck, Robles, & Turner, 2011; Markovikj, Gievska, Kosinski, & Stillwell, 2013; Staiano et al., 2012), personal values (Chen, Hsieh, Mahmud, & Nichols, 2014; Gou, Zhou, & Yang, 2014), mental health issues like depression and PTSD (De Choudhury, Gamon, Counts, & Horvitz, 2013; Coppersmith, Harman, & Dredze 2014; De Choudhury et al., 2013), and even if an account is spam (Golbeck, 2015). These algorithms work on many platforms—Twitter, Facebook, Instagram, Pinterest, and so forth—and using many different types of data, including profile photos, Facebook likes, profile data, text the user has written, and more.

While exciting from a scientific perspective and useful in some personalized interface applications (discussed more below), this research also raises privacy concerns. If personal and potentially sensitive information about people can be automatically detected, even if users have explicitly chosen to keep that information private, many personal privacy boundaries are eroded. That is even more troublesome when the resulting inferences can be sold, used to target ads, or even used in hiring decisions. How to handle these issues is a question that is very much being debated.

Understanding the World from Social Network Signals

The fact that so much of the world is online, sharing thoughts and commentary on everyday life means that, taken together, posts can paint a picture of what the world (or specific parts of it) is talking about and how they feel about it. This has spurred research that leverages social media to understand things about the world.

Predicting election outcomes has been of particular focus, and it has been analyzed in studies across Europe and North America (Ceron, Curini, Iacus, & Porro, 2014; Sang & Bos, 2012; Tumasjan, Sprenger, Sandner, & Welpe, 2010). Success has been mixed. While many project have been able to build models, they are not always transferrable to other elections or other contexts, so much work remains to be done in this area. Similar techniques have been used for making other future predictions, including predicting box office revenue from Twitter posts (Asur & Huberman, 2010) and the stock market (Bollen, Mao, & Zeng, 2011).

Social networks, and especially Twitter, have been studied as a source of breaking news and trending topics. HCI research has focused on tracking how individual events spread, like the death of Osama Bin Laden (Hu et al., 2012). While mass media outlets are popular within the social networking platforms, celebrities and other influential players often help spread the news. This is echoed in work that looked at the Boston Marathon Bombing and other terrorist events, and found official accounts and media outlets were most likely to gain attention, but initial reports often came from a variety of sources (Buntain, 2015).

A Phenomenon to Study Itself

How social networking platforms are used has become a topic to study in its own right. Early work simply described these sites and their dynamics, as the phenomenon was not well understood (Golbeck, 2007; Kaplan & Haenlein, 2010). Human-computer interaction research then moved on to study how different groups used social media platforms. Examples include university faculty (Moran, Seaman, & Tinti-Kane, 2011), the U.S. Congress (Golbeck, Grimes, & Rogers, 2010), students (Cheung, Chiu, & Lee, 2011), libraries (Aharony, 2012; Hendrix, Chiarella, Hasman, Murphy, & Zafron, 2009), and people with eating disorders (Mabe, Forney, & Keel, 2014).

That work continues, and research has also evolved to study particular uses of social media. That can be offline events organized through social media, like the Arab Spring (Gerbaudo, 2012; Lotan et al., 2011). However, the most prominent current topic in this area is that of hashtag movements. These are ways to build community and draw attention to issues with no real parallel in the offline world.

Research includes studies of #YesAllWomen (Rodino-Colocino, 2014), #occupywallstreet (Gleason, 2013), #gamergate (Massanari, 2015), and antisexism in science hashtags like #distractinglySexy (Golbeck, Ash, & Cabrera, 2016).

Human-computer interaction researchers are also studying how people use various platforms. On Instagram, research looks at photo content (Hu, Manikonda, & Kambhampati, 2014). On Pinterest, it's how people curate photos (Hall & Zarro, 2012; Zarro & Hall, 2012) and who uses it (Chang, Kumar, Gilbert, & Terveen, 2014). On Snapchat, it is the ephemeral nature of the interactions (Bayer, Ellison, Schoenebeck, & Falk, 2016). There are also comparisons between sites, such as how interaction varies between Snapchat and Facebook (Utz, Muscanell, & Khalid, 2015).

Understanding how people use technology is indeed a core question to much HCI research, and as social networking technology continues to advance and evolve, more work is certain to continue on these topics.

A Challenge for Information Visualization

Social media presents many unique challenges for information visualization (see Figure 34.1). Some of that comes from the type of data, including social graphs, and some from the size. As a major source of interesting Big Data, visualizing and analyzing data from social networking sites can be challenging simply because of the scale.

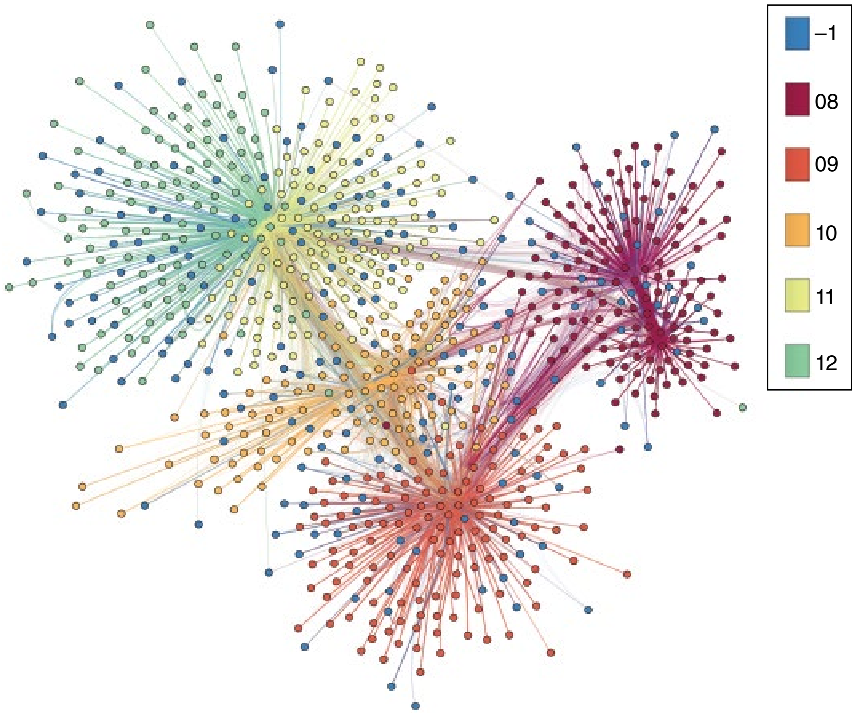


Figure 34.1 A sample social network visualization from Crnovrsanin, Muelder, Faris, Felmlee, & Ma (2014).

Visualizing the structure of the networks themselves is a challenge. A graph visualization can show useful insights with a few thousand nodes and the low tens of thousands of edges as a very upper limit. While that may sound large, online social networks are many orders of magnitude larger than this. A number of tools have been designed for this purpose, two of the most prominent being Gephi and NodeXL (Smith et al., 2009). Human-computer interaction researchers have also introduced techniques for hierarchically organizing graph data to simplify large networks, including both edge aggregation (Holten, 2006) and node hierarchies (Noel & Jajodia, 2004).

The data shared on social networking platforms also lends itself to visual analysis. On Instagram, daily rhythms appear in visual analysis of photos (Hochman & Schwartz, 2012). Word clouds pulled from social media text allow a visual inspection of a user's common topics (Viegas, Wattenberg, & Feinberg, 2009). Because so much social media data contains geospatial information, challenges have also arisen for combining social media and GIS data into visualizations (Sui & Goodchild, 2011).

Social Networks in Interfaces

Online social networks have impacted the interface side of HCI as well, particularly through the use of the data to drive personalized interface elements. Targeted advertising and content personalization are well known features that leverage social networking data.

Social recommender systems are among the most prominent HCI research areas that leverage information from social networks. Recommender systems in general personalize content based on a user's known preferences. This can be based on item similarity (i.e. find items that are similar to ones the user likes) or user similarity (i.e. find people who like similar things to the user and recommend other things those people like). Social recommenders leverage the fact that people tend to be friends with people who share their tastes, and use social connections to find items that might be of interest to a user.

Socially based recommendations tend to lead to increased interaction (Guy, 2015). These recommendations are particularly useful in taste-based domains (e.g. music or movies) (Groh, Birnkammerer, & Köllhofer, 2012), and they can often improve recommendation performance in cases where the user's taste lies well outside the norm (Golbeck & Hendler, 2006).

While traditional recommender systems tend to suggest items like books, movies, music, and other products that a user might want, friend recommenders suggest people a user might want to connect with socially. These are now ubiquitous interface elements in most online social networking sites. These may rely on similar algorithms to more classic recommender systems but even in the early days of online social networks, they often exploited the graph structure to gain insights (Lo & Lin, 2006; Moricz, Dosbayev, & Berlyant, 2010). Newer approaches to friend recommendation are leveraging the rich data sources available from mobile devices and other lifestyle trackers to find potential contacts for users in social environments (Wang, Liao, Cao, Qi, & Wang, 2015).

Conclusion

Online social networks have had one of the most profound impacts on the way people interact with technology. They made creation and interaction on the Internet possible for hundreds of millions of people. Through their use, researchers now have new ways of understanding people and the world, a broad set of technologies that are interesting in and of themselves, and a growing set of data that feeds interface features and personalization.

However, as social networks settle in as an important part of most people's lives, there are many new challenges ahead, including those that will face the HCI community. Privacy issues are already important and will only continue to become more critical. The analysis and use of social networking data can be useful, but it can also be used in ways that people would never approve of or consent to. Right now, users have limited power to control their personal data, and designing better interfaces, tools, and techniques for understanding the privacy implications of online sharing is important.

Online harassment is also a problem that has infiltrated most social networking spaces. A variety of techniques will be needed to combat this so that these online spaces can remain safe and effective places to interact.

Beyond these immediate issues, social networking is still wide open for analysis and for the application of existing theories to understand user behavior. The full spectrum of human behavior, thought, and interaction is now being recorded and shared every day in volumes that far exceed our processing abilities. Human-computer interaction researchers have a wealth of data and need only consider which aspects to study and which tools they will bring to bear in order to continue producing interesting insights about how people interact with and through this technology.

References

Aharony, N. (2012). Facebook use in libraries: an exploratory analysis. In *Aslib Proceedings*, *64*, 358–372.

Asur, S., & Huberman, B. A. (2010). Predicting the future with social media. In *2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT)* (vol. 1, pp. 492–499). New York, NY: ACM.

Avesani, P., Massa, P., & Tiella, R. (2005). A trust-enhanced recommender system application: Moleskiing. In *Proceedings of the 2005 ACM Symposium on Applied Computing* (pp. 1589–1593). New York, NY: ACM.

Back, M. D., Stopfer, J. M., Vazire, S., Gaddis, S., Schmukle, S. C., Egloff, B., & Gosling, S. D. (2010). Facebook profiles reflect actual personality, not self-idealization. *Psychological Science*, *21*(3), 372–374.

Bayer, J. B., Ellison, N. B., Schoenebeck, S. Y., & Falk, E. B. (2016). Sharing the small moments: ephemeral social interaction on snapchat. *Information, Communication and Society*, *19*(7), 956–977.

Bollen, J., Mao, H., & Zeng, X. (2011). Twitter mood predicts the stock market. *Journal of Computational Science*, *2*(1), 1–8.

Buntain, C. (2015). Discovering credible events in near real time from social media streams. In *Proceedings of the 24th International Conference on World Wide Web* (pp. 481–485). doi:10.1145/2740908.2741752

Ceron, A., Curini, L., Iacus, S. M., & Porro, G. (2014). Every tweet counts? How sentiment analysis of social media can improve our knowledge of citizens political preferences with an application to Italy and France. *New Media and Society*, *16*(2), 340–358.

Chang, S., Kumar, V., Gilbert, E., & Terveen, L. G. (2014). Specialization, homophily, and gender in a social curation site: findings from pinterest. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 674–686). New York, NY: ACM.

Chen, J., Hsieh, G., Mahmud, J. U., & Nichols, J. (2014). Understanding individuals' personal values from social media word use. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing* (pp. 405–414). New York, NY: ACM. doi:10.1145/2531602.2531608

Cheung, C. M., Chiu, P.-Y., & Lee, M. K. (2011). Online social networks: Why do students use Facebook? *Computers in Human Behavior*, *27*(4), 1337–1343.

Coppersmith, G., Harman, C., & Dredze, M. (2014). Measuring post traumatic stress disorder in Twitter. In *Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media* (pp. 579–582). Palo Alto, CA: AAAI.

Crnovrsanin, T., Muelder, C. W., Faris, R., Felmlee, D., & Ma, K. L. (2014). Visualization techniques for categorical analysis of social networks with multiple edge sets. *Social Networks*, *37*, 56–64.

De Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013). Predicting depression via social media. In *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media* (pp. 128–137). doi:10.1109/WI-IAT.2015.166

Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of facebook friends: social capital and college students use of online social network sites. *Journal of Computer-Mediated Communication*, *12*(4), 1143–1168.

Gerbaudo, P. (2012). *Tweets and the streets: Social media and contemporary activism*. London: Pluto Press.

Gilbert, E., & Karahalios, K. (2009). Predicting tie strength with social media. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 211–220). doi:10.1145/1518701.1518736

Gleason, B. (2013). # Occupy Wall Street: Exploring informal learning about a social movement on Twitter. *American Behavioral Scientist*, *57*(7), 966–982.

Golbeck, J. (2005). *Computing and applying trust in web-based social networks* (Unpublished PhD dissertation). University of Maryland, College Park, MD.

Golbeck, J. (2007). The dynamics of web-based social networks: Membership, relationships, and change. *First Monday*, *12*(11). doi:10.5210/fm.v12i11.2023

Golbeck, J. (2009). Trust and nuanced profile similarity in online social networks. *ACM Transactions on the Web (TWEB)*, *3*(4), 12.

Golbeck, J. (2015). Benford's law applies to online social networks. *PLOS One*, *10*(8), e0135169.

Golbeck, J., Ash, S., & Cabrera, N. (2017). Hashtags as online communities with social support: A study of anti-sexism-in-science hashtag movements. *First Monday*, *22*(9). doi:10.5210/fm.v22i9.

Golbeck, J., Grimes, J. M., & Rogers, A. (2010). Twitter use by the us congress. *Journal of the American Society for Information Science and Technology*, *61*(8), 1612–1621.

Golbeck, J., & Hansen, D. (2014). A method for computing political preference among twitter followers. *Social Networks*, *36*, 177–184.

Golbeck, J., & Hendler, J. (2006). Filmtrust: Movie recommendations using trust in web-based social networks. *Proceedings of the IEEE Consumer Communications and Networking Conference*, *96*, 282–286.

Golbeck, J., Robles, C., & Turner, K. (2011). Predicting personality with social media. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems* (pp. 253–262). doi:10.1145/1979742.1979614

Gou, L., Zhou, M. X., & Yang, H. (2014). Knowme and shareme: Understanding automatically discovered personality traits from social media and user sharing preferences. In *Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems* (pp. 955–964). New York, NY: ACM. doi:10.1145/2556288.2557398

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, *78*(6), 1360–1380.

Groh, G., Birnkammerer, S., & Köllhofer, V. (2012). Social recommender systems. In J. J. Pazos Arias, A. Fernández Vilas & R. P. Díaz Redondo (Eds.), *Recommender systems for the social web* (pp. 3–42). Berlin: Springer.

Guy, I. (2015). Social recommender systems. In P. B. Kantor (Ed.), *Recommender systems handbook* (pp. 511–543). Berlin: Springer.

Hall, C., & Zarro, M. (2012). Social curation on the website pinterest. com. *Proceedings of the American Society for Information Science and Technology*, *49*(1), 1–9.

Hendrix, D., Chiarella, D., Hasman, L., Murphy, S., & Zafron, M. L. (2009). Use of facebook in academic health sciences libraries. *Journal of the Medical Library Association*, *97*(1), 44.

Hochman, N., & Schwartz, R. (2012). Visualizing Instagram: Tracing cultural visual rhythms. In *Proceedings of the Workshop on Social Media Visualization (SOCMEDVIS) in Conjunction with the Sixth International AAAI conference on Weblogs and Social Media (ICWSM–12)* (pp. 6–9). Palo Alto, CA: AAAI Publications.

Holten, D. (2006). Hierarchical edge bundles: Visualization of adjacency relations in hierarchical data. *IEEE Transactions on Visualization and Computer Graphics*, *12*(5), 741–748.

Hu, M., Liu, S., Wei, F., Wu, Y., Stasko, J., & Ma, K.-L. (2012). Breaking news on twitter. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2751–2754). New York, NY: ACM. doi:10.1145/2207676.2208672

Hu, Y., Manikonda, L., & Kambhampati, S. (2014). What we Instagram: A first analysis of Instagram photo content and user types. In *Proceedings of the Eighth International Conference on Weblogs and Social Media (ICWSM 2014)* (pp. 595–598). Palo Alto, CA: AAAI Press.

Jernigan, C., & Mistree, B. F. (2009). Gaydar: Facebook friendships expose sexual orientation. *First Monday*, *14*(10). Retrieved from http://www.firstmonday.dk/ojs/index.php/fm/article/view/2611/2302

Kaplan, A. M., & Haenlein, M. (2010). Users of the world, unite! the challenges and opportunities of social media. *Business Horizons*, *53*(1), 59–68.

Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, *110*(15), 5802–5805.

Lo, S., & Lin, C. (2006). Wmr—a graph-based algorithm for friend recommendation. In *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence* (pp. 121–128). Washington, DC: IEEE Computer Society.

Lotan, G., Graeff, E., Ananny, M., Gaffney, D., Pearce, I., & Boyd, D. (2011). The Arab spring| the revolutions were tweeted: Information flows during the 2011 Tunisian and Egyptian revolutions. *International Journal of Communication*, *5*, 31.

Massanari, A. (2015). # gamergate and the fappening: How Reddit's algorithm, governance, and culture support toxic technocultures. *New Media and Society*, *19*(3), 329–346.

Mabe, A. G., Forney, K. J., & Keel, P. K. (2014). Do you like my photo? Facebook use maintains eating disorder risk. *International Journal of Eating Disorders*, *47*(5), 516–523.

MacKenzie, I. S. (1992). Fitts' law as a research and design tool in human-computer interaction. *Human-Computer Interaction*, *7*(1), 91–139.

Markovikj, D., Gievska, S., Kosinski, M., & Stillwell, D. (2013). Mining Facebook data for predictive personality modeling. In *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media (ICWSM 2013)* (pp. 23–26). Palo Alto, CA: AAAI Press.

Moran, M., Seaman, J., & Tinti-Kane, H. (2011). *Teaching, learning, and sharing: How today's higher education faculty use social media*. Boston, MA: Pearson Learning Solutions and Babson Survey Research Group. Retrieved from http://files.eric.ed.gov/fulltext/ED535130.pdf

Moricz, M., Dosbayev, Y., & Berlyant, M. (2010). Pymk: Friend recommendation at Myspace. In *Proceedings of the 2010 ACM SIGMOD International Conference on Management of Data* (pp. 999–1002). New York, NY: ACM.

Noel, S., & Jajodia, S. (2004). Managing attack graph complexity through visual hierarchical aggregation. In *Proceedings of the 2004 ACM Workshop on Visualization and Data Mining for Computer Security* (pp. 109–118). New York, NY: ACM.

Pfeil, U., Arjan, R., & Zaphiris, P. (2009). Age differences in online social networking—a study of user profiles and the social capital divide among teenagers and older users in Myspace. *Computers in Human Behavior*, *25*(3), 643–654.

Preece, J. (2000). *Online communities: Designing usability and supporting sociability*. New York, NY: John Wiley & Sons, Inc.

Rodino-Colocino, M. (2014). # yesallwomen: Intersectional mobilization against sexual assault is radical (again). *Feminist Media Studies*, *14*(6), 1113–1115.

Sang, E. T. K., & Bos, J. (2012). Predicting the 2011 Dutch senate election results with twitter. In *Proceedings of the Workshop on Semantic Analysis in Social Media* (pp. 53–60). Stroudsburg, PA: Association for Computational Linguistics.

Smith, M. A., Shneiderman, B., Milic-Frayling, N., Mendes Rodrigues, E., Barash, V., Dunne, C.,…Gleave, E. (2009). Analyzing (social media) networks with nodexl. In *Proceedings of the Fourth International Conference on Communities and Technologies* (pp. 255–264). doi:10.1145/1556460.1556497

Staiano, J., Lepri, B., Aharony, N., Pianesi, F., Sebe, N., & Pentland, A. (2012). Friends don't lie: Inferring personality traits from social network structure. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing* (pp. 321–330). New York, NY: ACM. doi:10.1145/2370216.2370266

Steinfield, C., Ellison, N. B., & Lampe, C. (2008). Social capital, self-esteem, and use of online social network sites: A longitudinal analysis. *Journal of Applied Developmental Psychology*, *29*(6), 434–445.

Sui, D., & Goodchild, M. (2011). The convergence of gis and social media: challenges for giscience. *International Journal of Geographical Information Science*, *25*(11), 1737–1748.

Tumasjan, A., Sprenger, T. O., Sandner, P. G., & Welpe, I. M. (2010). Predicting elections with twitter: What 140 characters reveal about political sentiment. *Proceedings of the International Conference on Weblogs and Social Media*, *10*(1), 178–185.

Utz, S., Muscanell, N., & Khalid, C. (2015). Snapchat elicits more jealousy than facebook: a comparison of snapchat and facebook use. *Cyberpsychology, Behavior, and Social Networking*, *18*(3), 141–146.

Viegas, F. B., Wattenberg, M., & Feinberg, J. (2009). Participatory visualization with wordle. *IEEE Transactions on Visualization and Computer Graphics*, *15*(6), 1137–1144.

Wang, Z., Liao, J., Cao, Q., Qi, H., & Wang, Z. (2015). Friendbook: A semantic-based friend recommendation system for social networks. *IEEE Transactions on Mobile Computing*, *14*(3), 538–551.

Zarro, M., & Hall, C. (2012). Pinterest: Social collecting for# linking# using# sharing. In *Proceedings of the 12th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 417–418). New York, NY: ACM.

Ziegler, C.-N., & Lausen, G. (2005). Propagation models for trust and distrust in social networks. *Information Systems Frontiers*, *7*(4–5), 337–358.

35

Measuring Attitudes Online: Social Interaction and the Internet

Mary Joyce and Eve Griffin

Introduction

The discipline of human-computer interaction (HCI) has traditionally concerned itself with how users interact with technology, focusing in particular on evaluating how usable these technologies are. Users have therefore played a central role in the design and evaluation of technologies and systems that we use in our everyday lives. Previously, HCI has focused on technology as a tool—where systems are designed to complete specified tasks. However, technologies have evolved to support social interactions, where the technology acts as a social space. Computer networks, now more than ever, are being used to connect people to people. Growing from a set of protocols developed in 1989 by Berners-Lee to allow a research team to share data more efficiently, the World Wide Web has become not just a tool for researchers, scientists and academics, but an accessible space for all people. These networks allow people to create a range of social spaces in which to meet and interact with one another.

McNamara & Kirakowski (2006) have suggested the use of a three factor model for the evaluation of interactions between individuals and technology. This model suggests that functionality, usability and experience are three aspects of technology use that should be considered during evaluation. In particular, experience refers to the wider relationship between a user and technology—addressing more subjective aspects to the interaction between these two elements. This and other models of HCI research acknowledge that the users' perceptions, attitudes, and feelings towards technology are the key to understanding its broader impact.

In this chapter, we are primarily concerned with user interaction via the Internet, where the technology is used as a medium, specifically in relation to their attitudes. Psychology can help us to understand a lot about how users engage with technology and in turn how it is perceived, ultimately giving us a better understanding of the role of technology in society. In the context of HCI, aspects of usability such as attitude measurement can help us to understand how and why individuals interact with technology, and this can help to ensure that the technologies we design and develop best support the activities we perform with these systems. To this end, the objectives

The Wiley Handbook of Human Computer Interaction, Volume 2, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

of this chapter are to illustrate how measurement of psychological constructs (in this case, measurement of attitudes) is an essential part of understanding how we engage with and interact with technology (the Internet). We use two case studies to demonstrate this; the first presents an example of an attitude scale that was developed to better understand the factors associated with engaging in social friendships online; the second outlines information about a general Internet attitudes scale that can be used to understand how people feel about engaging with the Internet.

The Internet and Social Interaction

Uses of the Internet

A recent study of adult Internet users in the United Kingdom found that 87% of respondents access the Internet, spending an average of 22 hours online per week (Ofcom, 2016). Popular online activities by users include communication, general surfing/browsing, and entertainment (including gaming and watching videos online). Socially-oriented activities on the Internet have progressively increased in recent years, with more people meeting friends and maintaining personal relationships online (Anderson, 2005). For example, the recent Ofcom (2016) study found that 93% of Internet users go online to send or receive mail, 76% access social networking sites, and 78% use the Internet for instant messaging. Our reasons for accessing the Internet are shifting from purely task- and goal-oriented activities (e.g. information seeking) to more socially oriented activities spurred on by the popularity of social networking sites such as Facebook and Twitter. In 2015 alone, Facebook and Twitter reported 1.4 billion and 300 million active users respectively, which confirms reports that social activity is the most popular reason for accessing the Internet (e.g. Cummings, Butler, & Kraut, 2002). Online gaming has also become increasing popular with almost 5 in 10 Internet users playing games online and 10% considering themselves to be “gamers” (Duggan, 2015).

Individual Differences and the Internet

As with other topics in HCI, social interaction online is mediated by individual differences such as gender, age and level of experience/confidence with engaging with this technology.

Gender In an attempt to better understand how males and females interact with technology, many studies have investigated the role of gender with regard to technology use. There has been a range of studies exploring differences among men and women and their interactions with the Internet, with a range of contradictory results emerging. Such conflicting results are not surprising given that many methodological issues have been identified in studies including: lack of consistent measurement methods to evaluate relevant constructs; lack of theoretical frameworks to guide how best to measure the construct under investigation; and sampling issues as a result of arbitrarily limited populations. These issues have resulted in insecure data from which it is difficult to draw reliable conclusions. Nonetheless, these studies *do* express a concern that there might well be a difference between men and women in their reactions to technology that is seen as predominantly governed by men. For example, studies carried out by Joiner et al. (2005), Jones, Johnson-Yale, Millermaier, and Perez (2009), and Helsper (2010) found that males used the Internet more often than females. More recently however, it would seem that this gender gap may be decreasing with the overall estimated weekly volume of Internet use reportedly similar for men and women (Ofcom, 2016). While the gender gap with regard to time spent online has decreased in recent years, there are still differences in the ways in which males and females use the Internet. Men are more likely than women to have higher use in the workplace / place of education and are more likely to use the Internet for task-oriented activities (e.g. work, information searching) while women are more likely to use the Internet for social media (Muscanell & Guadagno, 2012). Kimbrough, Guagagno, Muscanell and Dill (2013) also found that not only were women more likely to use social networking sites to maintain relationships—they were also more likely to integrate text-based / mediated communication in this practice.

Age The Internet is a relatively new phenomenon so it is possible that a division exists between individuals who have grown up immersed in an Internet environment and those to whom the Internet was introduced at a later stage in life. Bennett, Maton, and Kervin (2008) refer to the generation born between the 1980s and mid-1990s as "digital natives," a term first coined by Marc Prensky (2001) to describe the familiarity with, and dependence on, information and communications technology in younger generations. Similar to the notion of "digital natives", Howe and Strauss (1990) refer to individuals born between 1982 and c.2002 as "millennials". Both terms refer to the young individuals who are characterized by their upbringing in an immersive technological environment. However, those of a later generation, who Prensky (2001) refers to as "digital immigrants", have had to adapt and integrate technology in their lives. Researchers such as Van Deursen, Van Dijk and Peters (2011) suggest that younger generations are particularly skilled users of the Internet as they have had exposure to the Internet throughout their entire life. They refer to research carried out by de Haan and Rijken (see Van Deursen et al., 2011) who suggest that seniors did not have the opportunity to acquaint themselves with the Internet at school, and thus lack the same level of use and ownership of Internet skills. A study by Zhang (2005) which investigated age differences in Internet attitudes amongst employees in a business environment found that younger employees (<20 years old) felt that the Internet was more useful than any of the older age groups. The Ofcom (2016) report also found that younger Internet users (16–24 year olds) have the highest weekly volume of use of all Internet users. Despite younger age groups being associated with increased computer and Internet use, age has been regarded as a strong predictor of performance in terms of computer-related tasks, irrespective of experience. While there are conflicting reports regarding age differences with regard to Internet abilities, in general, research is lacking on age as a factor for consideration in HCI research; possibly because work in this area has been primarily conducted on student populations.

The importance of measuring individual differences

From the evidence presented thus far, we know that there are individual differences present in *how* people interact with the Internet. However, being aware of trends of online interaction, in particular social online interaction, does not provide us with enough information to determine *why* such trends occur. It is understandable that researchers, policy makers and educators want to know the reasons for online behavior. It is for reasons such as these that HCI researchers and psychologists look to constructs such as attitude and self-efficacy to help them understand why such trends in Internet use occur. Turning our attention to the construct of attitude in particular, considering that attitudes are thought to consist of three components (affect, behavioral intention and cognition), they have the potential to provide HCI researchers with important information about how individuals feel about and plan their future interactions with technology. With the constant increasing use of the Internet worldwide, it is important that people's attitudes are evaluated in order to maximize the potential of this ever-growing device.

Measuring Individual Differences in HCI: Attitude Measurement

Measurement in HCI

Standardized tests are widely used in the study of HCI or for testing hypotheses and one of the main advantages of their use in user testing is their ability to elicit subjective responses from a user. These scales are applied to measure, for example, the perceived usability of a software system (e.g. SUMI, SUS) or a website (e.g. WAMMI), or more subjective measurement of a person's emotional response to a piece of technology (e.g. Computer Anxiety Rating Scale). As outlined in this chapter, using scale instruments to measure attitudes is a frequently used technique in social psychology. Psychometric questionnaires, while resource intensive and time consuming to construct, have a longer "shelf life", and when used on a long-term basis the benefits outweigh the costs (Kirakowski & Corbett, 1990; McNamara & Kirakowski, 2011). Psychometric questionnaires differ from survey-based designs in that they are built upon previous research and they have demonstrated reliability and validity. A more detailed account of survey and scale development methods can be found in Bryman (2016) and Oppenheim (1992). The benefit of using such scales in HCI research allows for flexibility in study design. These scales may be used in traditional experimental settings or as online tools. Using the Internet as a research tool holds many advantages for data collection including speed, convenience, low costs, and anonymity (Miller, Johnston, Dunn, Fry & Degenhardt, 2010). Research shows that using the Internet to collect survey or scale responses may be successfully used to access specific populations that may be hidden or difficult to access (Coomber, 1997; Miller et al., 2010). Studies focusing on anonymity effects can imply that responses provided via the Internet are more honest and accurate than those given in more traditional face-to-face interviews (Davis, 1999; Joinson, 1999). It is also thought that people are more likely to complete open-ended questions when presented in an online format (Norman, 2008).

Statistical techniques

The most common statistical method employed to develop standardized measurement scales is factor analysis. This analysis works on the assumption that emerging factors reflect underlying processes that have created the correlations between individual scale items. This approach begins with the compilation of a large list of items that is reduced (using this statistical technique) to a smaller set of items that effectively represents the concept under investigation. The initial item list is piloted and from these scores, factors are derived. As a result of the initial factor analysis, items are added and deleted accordingly and the process continues until a scale with numerous items forming several factors representing the measurement area is formed (Tabachnick & Fidell, 2013). Exploratory factor analysis is used to describe and summarize the data by grouping correlated variables, and is used in the early stages of research. Confirmatory factor analysis is used at later stages to test the emergent structure of a scale—where hypotheses about which factors particular items should load on, as well as the number of factors that should emerge, are applied. See Tabachnick and Fidell (2013) and Bandalos and Finney (2010) for more detailed information on these techniques.

Background to attitudes

Agreement on a definition of attitudes has initiated much debate amongst psychologists. Numerous definitions of attitude have been proposed since the beginning of attitude research in the early 20th century. Fishbein and Ajzen's (1975) description of an attitude has held significance in social psychology, where they define an attitude as "a person's feelings toward and evaluation of some object, person, issue, or event" (p. 12). Note the emphases on *feelings toward* and *evaluation* in their definition. Some years later, Eagly and Chaiken (2007) defined an attitude as "a psychological tendency that is expressed by evaluating a particular entity with some degree of favor or disfavor" (p. 1). Keeping the outlined definitions in mind, this research emphasizes both the *feeling* and *evaluative* elements of attitudes. Various models of attitudes have been proposed since the beginning of attitude research. These models have close connections to the definitions of attitudes that have been proposed throughout the duration of attitude research. One, two and three component models of attitude have been considered over many years of research. The three-component model is the most favored model in attitude research to date. The three components of attitude are described as *affect*—an emotion that charges the idea; a feeling that may be good or bad when thinking about the attitude object; *behavioral intention*—the individual's predisposition to action in regard to the attitude object; and *cognition*—the beliefs and ideas a person has about the attitude object.

Measuring attitudes

Attitudes cannot be directly observed so it is understandable that they cannot be *measured* directly. Accordingly, the most obvious way to measure something in a person's mind is to ask the individual themselves to describe their own attitudes. However, this often generates problems for the researcher because people's words do not always correspond to their inner attitudes (Forsyth, 1987). In fact, J. Kirakowski (personal communication, 2012) suggests that the actual problem is that individuals have subjective interpretations of language to describe their personal feelings and beliefs. This results in discrepancies and individual differences regarding the verbal documentation of an attitude. DeVellis (2003) outlines that the reason we develop scales is to measure phenomena that we believe exist in our theoretical understanding of the

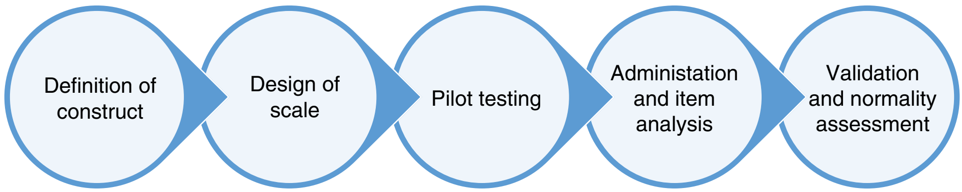


Figure 35.1 Five steps to measuring attitudes. *Source:* Adapted from Spector (1992).

world but that cannot be assessed directly. Generally, methods of attitude measurement are based on the assumption that attitudes can be measured by the opinions or beliefs of people regarding the attitude objects (Stahlberg & Frey, 1996). Likert's method of summated ratings is the most commonly used scale for measuring attitudes due to its ease of construction and interpretation, and its statistical tractability. In addition, Likert's method has contributed significantly to modern questionnaires that assume that the attitude being measured may have underlying dimensions (Hogg & Vaughan, 2011). Furthermore, the Likert method is commonly used in attitude scales due to its ease of interpretation by those who may not be familiar with psychometric methodologies. Likert scales are typically constructed using five positions for measurement of agreement (Hogg & Vaughan, 2011). Certainly in computer and Internet attitude research, Likert scales using five position statements (usually strongly disagree to strongly agree) are the prevailing practice.

DeVellis (2003) and Spector (1992) both present guidelines for the development of measurement scales. While DeVellis presents eight guidelines, which he outlines as "a set of specific guidelines that investigators can use in developing measurement scales" (p. 60), Spector lists five major steps with specific reference to the construction of a summated rating scale. As Spector's steps are specifically tailored for the construction of a summated rating scale and are presented in a more concise manner than those outlined by DeVellis, Spector's multistep process is outlined here. The five steps as outlined by Spector (1992) are shown in Figure 35.1. It should be noted that while DeVellis' (2003) guidelines pertain to the development of measurement scales in general, and Spector's guidelines are specific to a summated rating scale, the steps in each process are closely related. Each of DeVellis' eight steps is incorporated within Spector's five-step process of scale development.

Keeping the above outlined principles for scale development and measurement in HCI research in mind, we now present two case studies that document the development of scales to measure attitudes to Internet relationships and attitudes towards the Internet. We then go on to discuss the applications and uses of these scales in various research contexts.

Case Study 1: Development of the Attitude to Internet Relationships Scale (AIRS)

Background and overview

The development of the Attitude to Internet Relationships Scale (AIRS) emerged as a result of wanting to explore and uncover the main themes or factors that are important aspects of how and why people communicate online. As we have seen already in this chapter, the adoption of technology that mediates social communication has progressed rapidly over the past few years. However, no known study has developed a validated instrument that specifically assesses individual attitudes towards online relationships. The scale is not limited to *romantic* relationships; however some of the scale items are relevant to romantic relationships. To date, no known study has developed an instrument that specifically assesses how people perceive the formation and maintenance of online relationships. Instead of simply looking at *amount* of Internet use that previous studies have tended to focus on, this study will investigate relationships between different types of online activities and attitudes to online relationship formation.

The finalized scale contains 32 items, and has demonstrated excellent reliability and validity. Principal components analysis produced three subscales: *Interactivity; Contexts of CMC;* and *Online Commitment*. Statistical analysis of a large sample of respondents (n = 604) found that people who use the Internet for social activities hold more positive attitudes to online relationships than those who do not engage in such activities, and differences were found between males and females in terms of the way in which they use the Internet for social communication. Furthermore, it is theorized that the scale highlights the salient factors influencing people's motives for engaging in online communication and relationships.

Methodology

The scale was developed over three phases of data collection and analysis. An item pool of 64 generated items was used in the pilot phase. All items were attitudinal statements, both positively and negatively worded, and examined the extent to which the participants agreed with each one. Items were scored on a five-point Likert scale and the response anchors were: 1 = strongly agree; 2 = agree; 3 = no opinion; 4 = disagree; 5 = strongly disagree. The questionnaire was preceded by a section collecting demographic information about participants, and followed by a section where participants were encouraged to give an account of any Internet interactions and relationships they had experienced. High scores on the scale would indicate more positive attitude levels toward online encounters and interactions.

Across three stages of data collection using both online and paper formats, the item pool was reduced to a small set of items that would adequately measure the construct under investigation. Additionally, patterns of commonality among items (subscales) could also be identified in this stage. A total of 528 responses were gathered across these three phases of data collection.

Analysis

The final analysis produced the best solution of three factors. The final version of the scale consists of 32 items. The three emerging factors had Eigenvalues of 6.8, 3.6, and 2.9, respectively. Together, the factors accounted for a total of 38.7% of the variance in item responses (Factor 1 = 18%; Factor 2 = 9%; Factor 3 = 8%).

Reliability

The reliability of each subscale was assessed to ensure that the construct being measured was appropriately reflected (internal consistency). All Cronbach's alpha values were above 0.6 (see Table 35.1).

Table 35.1 Factorial structure of Attitude to Internet Relationships Scale (AIRS) with item examples.

	Interactivity (n = 11) $\alpha = 0.62$	*Contexts of CMC (n = 11)* $\alpha = 0.79$	*Online commitment (n = 11)* $\alpha = 0.81$
Item examples	It's easy to be deceived by someone over the Internet	I find it exciting getting to know people through the Internet	I tend to have very low expectations of people I talk to over the Internet
	Certain aspects of someone's personality can be easily hidden on the Internet	I enjoy role playing on the Internet	It is easy to end Internet relationships
	I can present myself in different ways on the Internet	I like the way that the Internet reduces physical stereotypes and prejudices	There are fewer emotional consequences with Internet relationships

Factorial structure of the AIRS

Reducing the factors to a three-way solution in the final rotation not only reduces the number of items in each subscale but also creates more comprehensive and conceptual constructs. The final three factors group more specific aspects of computer-mediated communication into three broad characteristics that relate to social communication in online settings (see Table 35.1). In naming the factors, the intention was to produce names that would clearly and adequately communicate the nature of the underlying construct. Semantic patterns between the items on each factor were investigated, and the names attributed to the three factors were influenced by the extant literature. To this end, the three factors (or subscales) were named as follows: *Interactivity* (11 items), *Contexts of CMC* (11 items), and *Online Commitment* (10 items).

The emergent factorial structure of the questionnaire is important to consider as it appears to reflect underlying patterns in the surrounding literature of relationships in cyberspace. The first subscale relates to *interactivity*. Here, interactivity refers to the extent that people think they are in control of the Internet situation and can give or withhold private information about themselves. This subscale seeks to measure the "different feelings, experiences, or perceptions of interactivity of different levels or intensity" that a user may subjectively hold (Rafaeli & Ariel, 2007, p. 82). The second subscale, *Contexts of CMC*, refers to the motivations and contexts for interacting via computer-mediated communication. This incorporates attitudes to the various activities that people engage with online. It also refers to people's attitudes towards the meaning and quality of relationships formed online. The final subscale, *Online Commitment*, encompasses people's attitudes to the emotional value placed on online relationships. The attributed meaning, importance and values of relationships that have been formed online are incorporated in this subscale. Whitty and Gavin's (2001) qualitative exploration of online communication identified a theme relating to "online commitment," and so their term is adopted here.

Validation

The validation of the AIRS consisted of two stages: (a) conducting a confirmatory factor analysis and (b) hypothesis testing. For this, a sample of 604 Internet users were gathered. The sample consisted of 398 females and 204 males. 81% of the respondents were in the age bracket 16–25 years. The confirmatory factor analysis was performed to test both the concurrent and discriminant validity of the three-factor scale. The goodness-of-fit index score for this analysis was 0.91, indicating that the model accounts for 91% of the variance and covariance of the variables, and furthermore supports the hypothesized three-factor structure of the scale, demonstrating its concurrent validity. The interfactor correlations were all low. The correlation between Interactivity and Contexts of CMC was 0.15; the correlation between Interactivity and Online Commitment was 0.37; finally the correlation between Contexts of CMC and Online Commitment was 0.13. These low intercorrelations demonstrate good discriminant validity for the scale.

The second part of the validation process involved testing a number of hypotheses, using the finalized scale. Studies that have investigated attitudes to online relationships among groups have generally reached consensus that those who engage in online communication and relationships hold more positive attitudes to online communication than those who do not (e.g. Donn & Sherman, 2002; Wildermuth, 2004).

Group differences

Respondents were categorized based on Weiser's (2001) dimensions of Internet use (socioaffective regulation and goods-and-information acquisition). An independent subjects t-test found a significant difference between the two groups in terms of their total scores on the scale $t(602) = 3.12$; $p < 0.005$. Thus, it would appear that the respondents who used the Internet for social activities such as chat rooms and social networking sites held more positive attitudes towards online relationships than those who engage in task-oriented online activities. There was a significant difference between those who use social networking sites and chat rooms and those who do not on the combined dependent variables [$F(3, 600) = 3.86$, $p = 0.009$] (see Table 35.2).

A two-way between-groups ANOVA was conducted to explore the impact of gender and age on the total scores of this scale. There was no significant main effect for age or gender [$F(1, 594) = 0.09$, $p = 0.77$]. Therefore, while the mean scores across the age groups appear different—the attitudinal scores decrease with age—they are not statistically different.

Conclusion

The results of the multivariate analysis conducted on the final Internet sample data found that, in general, people who engage in Web-based social activities (social networking and chat rooms) hold more positive attitudes towards online relationships than those who do not engage in such activities. Gender and age did not seem to impact on the scores. However, gender patterns in relation to Internet usage emerged during the analysis. It was found that while there was no difference between males and females in terms of their engagement with social activities online, there does seem to be a gender difference in the types of platforms used for these activities; a greater

Table 35.2 Group difference on AIRS scores.

			Total scale	*Subscale 1*	*Subscale 2*	*Subscale 3*
		N	*Mean (SD)*	*Mean (SD)*	*Mean (SD)*	*Mean (SD)*
Group	All	604	60.7 (18.2)	60.7 (18.2)	60.7 (18.2)	60.7 (18.2)
	Social	533	61.6 (18.1)	20.8 (4.8)	21.1 (8.5)	19.7 (9.4)
	Functional	71	54.3 (17.9)	19.5 (5.7)	19.0 (7.7)	15.8 (8.2)
Gender	Male	204	62.2 (17.8)	21.2 (5.1)	21.8 (8.5)	19.1 (9.2)
	Female	398	59.9 (18.4)	20.4 (4.9)	20.3 (8.3)	19.3 (9.5)
Age	16–25	486	60.7 (18.1)	20.6 (4.9)	21.0 (8.2)	19.2 (9.4)
	26–39	93	62.9 (18.7)	21.9 (5.0)	20.6 (9.1)	20.3 (9.5)
	40–65	22	52.7 (17.6)	18.2 (5.6)	17.4 (8.0)	17.2 (8.7)

percentage of males accessed chat rooms, while almost 60% of females used social networking sites, compared to only 27% of males.

Case Study 2: Development of the General Internet Attitude Scale (GIAS) Scale

Background and overview

The development of the General Internet Attitude Scale (GIAS) resulted from observations of individuals' behavior when completing online tasks for a website evaluation. Variation in individuals' reactions and responses to difficult tasks were observed; yet examination of the literature yielded surprisingly few explanations in any quantitative sense for such discrepancies in responses. In previous attempts at developing scales to explore such reactions and behaviors, there were methodological issues resulting from the failure to follow conventional methodologies for the development of scales (a detailed critique of previous research on attitude can be found in Joyce & Kirakowski, 2013). There was a need for a standardized psychometric scale, recognizable as such from well-established social psychological theory, which would measure attitude in an Internet environment. This case study thus focuses on the development of a scale that measures Internet attitudes.

Methodology

The General Internet Attitude Scale is a quantitative measures of individuals' attitudes towards the Internet, based on the well-known three-component model of attitude in social psychological theory. The three component model of attitude was the starting point for GIAS. It initially consisted of items relating to *affect* (feelings, likes / dislikes about the Internet), *behavioral intention* (the intention to act a certain way on the Internet), and *cognition* (beliefs and cognitions of individuals about the Internet). Four existing questionnaires for the measurement of Internet attitudes (Durndell & Haag, 2002; Tsai, Lin & Tsai, 2001; Weiser, 2000; Zhang, 2007) were consulted when generating items for the GIAS. These four questionnaires were chosen as the basis for the item pool as these studies exemplified the best, and most recent attempts at creating items depicting Internet attitudes. The creation of GIAS Ver.1 proceeded in two redactive stages. The first stage of redaction was to address previous methodological issues inherent in previous Internet attitude scale development. An initial item pool of 97 statements was used in the first stage of scale development. Statements that addressed specific uses of the Internet were removed as they simply describe patterns of use that may reflect individual circumstances and do not represent an underlying attitude. Examples of these statements include "I like to use the Internet to communicate with my friends" or "I use the Internet regularly throughout school." Other problematic items were statements that referred to feelings of confidence (a construct more analogous to self-efficacy) with using the Internet, for example "I feel confident using an Internet browser" or "I feel confident discussing questions with others through the Internet". The second stage of redaction was to apply the modal theoretical framework of attitudes to each item in the final item pool. Each remaining item was examined to identify whether or not the statement represented (or could represent) one of the three components of an attitude: affect, behavioral intention, or cognition. If a statement could be reworded to represent one of the components of an attitude, then this was done. If however, a statement did not in any way represent one of the components of an attitude, it was deleted.

The scale was developed over three phases of data collection and analysis. An item pool of 27 items was used in the pilot phase. All items were attitudinal statements, both positively and negatively worded, and examined the extent to which the participants agreed with each one. Items were scored on a five-point Likert scale and the response anchors were: 1 = strongly disagree; 2 = disagree; 3 = no opinion; 4 = agree; 5 = strongly agree. The questionnaire was preceded by a section collecting demographic information about participants. Across the three stages of data collection using both online and paper formats, the item pool was reduced to a small set of items that would adequately measure the construct that the subscale represented. A total of 1,777 responses were gathered across these three phrases of data collection (see Joyce & Kirakowski, 2015 for a detailed account of the GIAS development).

Confirmatory factor analysis

A confirmatory factor analysis ($N = 841$) confirmed the existing four factors, although it suggested that a number of items be removed. The final model achieved excellent model fit, resulting in the final version (GIAS Ver.4) of the scale that contains 21 statements and four subscales (see Table 35.3).

Each of the subscales and the total scale (Cronbach's alpha = 0.85) achieved satisfactory reliability values.

Gender differences

Numerous studies (e.g. Weiser, 2000; Tsai et al., 2001) have investigated gender differences with technology in an effort to better understand how males and females interact with it. Conflicting evidence for the existence of gender differences has been reported in such studies. It is possible that such conflicting evidence results from the use of questionnaires in which the construct under investigation is not clearly established or psychometrically robust. An independent samples t-test was conducted on the mean attitude scores between males and females to investigate if there were any

Table 35.3 Factorial structure of General Internet Attitude Scale (GIAS) with item examples.

	Internet affect *(n = 9)* *α = 0.87*	*Internet exhilaration* *(n = 3)* *α = 0.76*	*Social benefit of the Internet* *(n = 6)* *α = 0.79*	*Internet detriment* *(n = 3)* *α = 0.67*
Item examples	I feel overwhelmed by the Internet The Internet makes me feel anxious I feel at ease using the internet	The thought of going on the Internet is exciting to me I would like to stay on the Internet for as long as I can The idea of going on the Internet gives me a thrill	The use of the Internet is enhancing our standard of living The Internet is responsible for many of the good things we enjoy The Internet makes a positive contribution towards society	Using the Internet is harmful to people The Internet is dehumanizing to society Using the Internet can cause health problems

statistically significant differences between the groups. It was found that there were no significant differences in total attitude scores between males and females. These results suggest that males and females hold similar attitudes towards the Internet. The findings of Lewis (2002) with the Post Study System Usability Questionnaire (PSSUQ) also support the finding of no effect of gender.

Age differences

Age group differences were investigated in this study for reasons outlined earlier. Participants were grouped according to their age as follows: <18 years; 18–24 years; 25–34 years; 35–44 years; 45–54 years; 55–64 years; 65 years +. A one-way between-groups analyses of variance was conducted to explore differences in total attitude scores across the seven age groups. There was a statistically significant difference in attitude scores for the seven age groups: $F(6, 834) = 3.14$, $p < 0.01$. The effect size was 0.22. Post hoc analyses indicated that the difference between the means of the <18 years group ($M = 3.49$, $SD = 0.47$) and 25–34 years group ($M = 3.95$, $SD = 0.49$) approached significance ($p = 0.06$). A steady decline in attitude scores was observed over the remainder of the age groups as age increases beyond the 25–34 years age group (see Figure 35.2).

Conclusion

While no gender differences were found in Internet attitude scores, the results of the analyses highlight some age differences. Participants aged between 25 and 34 years achieved the highest attitude scores suggesting that this age group hold the most positive attitudes towards the Internet. On the other hand, the youngest and

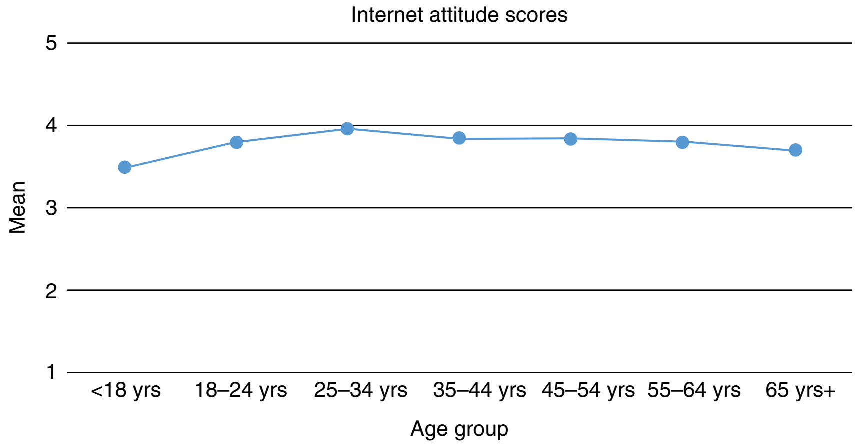


Figure 35.2 Internet attitude scores as a function of age.

oldest age groups (<18 years and 55 years+) achieved the lowest attitude scores of the sample indicating that they hold less favorable Internet attitudes than those of the other age groups. The number of participants in the <18 years age group was small however so findings for this age group should be interpreted with caution. The findings for the older age groups are in line with research that has investigated age differences in Internet use (e.g. Dutton & Blank, 2011; Ofcom, 2016).

Applications of the Scales

Previously developed attitude scales have been used in a variety of applied environments (e.g. HCI, e-learning, e-commerce, online relationships). For example, e-learning is becoming more prolific in our everyday lives as Internet connectivity and mobile devices continue to develop. While students rely more on the Internet to obtain information for their coursework than ever before, it is also commonplace for students to further their education through distance learning. Many courses also make use of forums and online classrooms within their institute's e-learning system to promote critical thinking, learning and peer-to-peer learning. As a result of this dependence on the Internet to assist learning, remote or otherwise, it is important to understand individuals' attitudes to the Internet to maximize the potential of such resources. Scales such as the GIAS and AIRS can be utilized to assess attitudes in such contexts. Such data obtained from individuals through these scales can quickly identify their attitudes towards the Internet, which then enables course developers to identify the best ways to maximize the delivery of the course and encourage student engagement.

The maintenance of relationships via online methods is commonplace in today's society through the ubiquitous use of smartphones and social media technology. Such tools are used to both support existing and often long-distance relationships and friendships, as well as nurturing new ones. Yet often such technologies are designed without consideration of the end users. The assessment of user attitudes and confidence with such tools is vital, yet often overlooked.

The GIAS has the potential to be applied to a wider range of Internet populations, not just limited to social users. An example of an expanding area relevant to this research is that of e-commerce. Many businesses now offer their products and services online to maximize the audience at which they can target. It is critical that businesses and service providers have interactive websites and interfaces that are pleasing, intuitive, and easy to engage with for users. Very often however, website developers test the usability of their website without gathering information from their target audience. The attitudes of users who evaluate the website is critical; evaluations of a website based on users who for example, may have a negative attitude about the Internet, may result in skewed results. A lot of research is done with relatively small, self-selected samples (Molich et al., 2010), so this is a real risk which can result in unnecessary changes being made to a website in a situation where the users' perception influences the results in a specious manner. Thus, the GIAS is an ideal tool to use in conjunction with usability studies that carry out such user testing.

With regard to the scales presented in this chapter, the factorial structure of the AIRS, which categorizes the salient categories for attitudinal measurement, is not to be taken as final, but merely as one that arose as a result of the scale's construction. Further work is being pursued to fully validate the scale and to assess its ultimate reliability and stability as an attitudinal instrument. However, in its current form, it is useful as an aid for gauging attitudes around these categories across and within groups of users and nonusers. It provides the opportunity to access large populations of people who engage in Internet-mediated communication across social spaces, at all levels of intimacy. Ideally, this scale can be implemented for use with more qualitative and exploratory methods to gain a deeper understanding into the views and opinions that people hold. The scale has already been used in an undergraduate research project (Corbett, 2011), which investigated the attitudes of stigmatized individuals to online relationships.

Conclusion

The purpose of this chapter was to provide some examples of how research in the field of HCI can assess the role of the user, by using existing validated scales or by developing instruments to measure a specific concept—in this case, attitudes. The first case study, the development of the Attitudes to Internet Relationships Scale (AIRS), demonstrated how a stable underlying three-factor structure emerged, with each factor referring to conceptual issues surrounding online communication. The initial validation stage of this questionnaire has shown that people who engage in online social interaction (via chat rooms and social networking sites) hold more positive attitudes towards online relationships than those who use the Internet for more task-oriented activities.

In this chapter, we argue that research should focus on the interaction between the different personalities of Internet users and the diverse components of Internet technology. McCarthy and Wright (2004) comment that technology should fit in with "a value system that treats communication and relationships as important" (p. 4). Human-computer interaction is no longer simply about finding a "fit" between person and technology with goals of increasing productivity, and the challenges become more diverse where goals and activities of people, their values, and the tools and mediums that shape their everyday lives are now of primary concern (Bannon, 2011). So, the focus of HCI research should not simply be on HCI, but should center around human activities mediated by technology, placing people at the core of the research process.

At a design level, Amichai-Hamburger (2002) argues that the Internet does not fulfil its potential as a medium for interaction. This he attributes to the lack of communication and sharing of knowledge between Internet designers and researchers. On one hand, Internet designers generally see Internet users as a homogenous group consisting of multiple instances of one stereotyped user profile and tend to ignore situational and individual differences. Conversely, many researchers may view the Internet as a single entity, and may ignore its richness and variety of services and the influences each service may have on communication.

New methods must be developed and adapted to account for the complexity of human nature, especially when applied to an Internet context. As the horizons of the Internet continually expand, our understanding of how people engage in this environment needs to keep pace. Only then can we inform the potentials and capabilities of these platforms as supports for human-to-human communication.

References

Amichai-Hamburger, Y. (2002). Internet and personality. *Computers in Human Behavior*, *18*(1), 1–10.

Anderson, T. L. (2005). Relationships among Internet attitudes, Internet use, romantic beliefs, and perceptions of online romantic relationships. *Cyberpsychology and Behavior*, *8*(6), 521–531.

Bandalos, D. L., & Finney, S. J. (2010). Factor analysis: Exploratory and confirmatory. In G. R. Hancock & R. O. Mueller (Eds.), *The reviewer's guide to quantitative methods in the social sciences* (pp. 93–114). New York, NY: Routledge.

Bannon, L. (2011). Reimagining HCI: Toward a more human-centered perspective. *Interactions*, *18*(4), 50–57.

Bennett, S., Maton, K., & Kervin, L. (2008). The "digital natives" debate: a critical review of the evidence. *British Journal of Educational Technology*, *39*(5), 775–786.

Bryman, A. (2016). *Social research methods* (5th ed.). New York, NY: Oxford University Press.

Coomber, R. (1997). Using the Internet for survey research. *Sociological Research Online*, *2*(2). Retrieved from http://www.socresonline.org.uk/2/2/2

Corbett, D. (2011). *Online communities: A new hope for stigmatized individuals* (Unpublished undergraduate thesis). University College Cork, Cork.

Cummings, J. N., Butler, B., & Kraut, R. (2002). The quality of online social relationships. *Communications of the ACM*, *45*(7), 103–108.

Davis, R. N. (1999). Web-based administration of a personality questionnaire: Comparison with traditional methods. *Behavior Research Methods, Instruments and Computers*, *31*(4), 572–577.

DeVellis, R. F. (2003). *Scale Development: theory and applications* (2nd ed.). Thousand Oaks, CA: Sage.

Donn, J. E., & Sherman, R. C. (2002). Attitudes and practices regarding the formation of romantic relationships on the Internet. *Cyberpsychology and Behavior*, *5*(2), 107–123.

Durndell, A. & Haag, Z. (2002). Computer self efficacy, computer anxiety, attitudes towards the Internet and reported experience with the Internet, by gender, in an East European sample. *Computers in Human Behavior*, *18*, 521–535.

Duggan, M. (2015). Gaming and gamers. Pew Research Center. Retrieved from www.pewinternet.org/2015/12/15/gaming-and-gamers/

Dutton, W. H., & Blank, G. (2011). *Next generation users: The Internet in Britain*. Oxford: Oxford Internet Institute, University of Oxford.

Eagly, A. H., & Chaiken, S. (2007). The advantages of an inclusive definition of attitude. *Social Cognition*, *25*(5), 582–602.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention and behaviour: An introduction to theory and research*. Reading, MA: Addison-Wesley.

Forsyth, D. R. (1987). *Social psychology*. Monterey, CA: Brooks / Cole.

Helsper, E. J. (2010). Gendered Internet use across generations and life stages. *Communication Research*, *37*(3), 352–374.

Hogg, M. A., & Vaughan, G. M. (2011). *Social psychology*. (6th ed.). Harlow: Pearson Education.

Howe, N., & Strauss, W. (1990). *Generations: The history of America's future, 1584 to 2069*. New York, NY: Morrow.

Joiner, R., Gavin, J., Duffield, J., Brosnan, M., Crook, C., Durndell, A.,…Lovatt, P. (2005). Gender, Internet identification, and Internet anxiety: Correlates of Internet use. *Cyberpsychology and Behavior*, *8*(4), 371–378.

Joinson, A. N. (1999). Social desirability, anonymity, and Internet-based questionnaires. *Behavior Research Methods, Instruments and Computers*, *31*(3), 433–438.

Jones, S., Johnson-Yale, C., Millermaier, S., & Perez, F. S. (2009). U.S. college students; Internet use: Race, gender and digital divides. *Journal of Computer-Mediated Communication*, *14*(2), 244–264.

Joyce, M. & Kirakowski, J. (2013). Development of a general Internet attitude scale. In A. Marcus (Ed.), *Design, user experience, and usability. Design philosophy, methods, and tools* (pp. 303–311). Berlin: Springer.

Joyce, M., & Kirakowski, J. (2015). Measuring attitudes towards the Internet: the General Internet Attitude Scale. *International Journal of Human-Computer Interaction*, *31*(8), 506–517.

Kimbrough, A. M., Guagagno, R. E., Muscanell, N. L., & Dill, J. (2013). Gender differences in mediated communication: Women connect more than do men. *Computers in Human Behavior*, *29*, 896–900.

Kirakowski, J. & Corbett, M. (1990). *Effective methodology for the study of HCI*. New York, NY: Elsevier.

Lewis, J. (2002). Psychometric evaluation of the PSSSQ using data from five years of usability studies. *International Journal of Human Computer Interaction 14*(3–4), 463–488.

McCarthy, J., & Wright, P. (2004). *Technology as experience*. London: MIT Press.

McNamara, N., & Kirakowski, J. (2006). Functionality, usability, and user experience: Three areas of concern. *Interactions*, *13*(6), 26–28.

McNamara, N. & Kirakowski, J. (2011). Measuring user satisfaction with electronic consumer products: The consumer products questionnaire. *International Journal of Human-Computer Studies*, *69*(6), 375–386.

Miller, P. G., Johnston, J., Dunn, M., Fry, C. L., & Degenhardt, L. (2010). Comparing probability and non-probability sampling methods in ecstasy research: Implications for the Internet as a research tool. *Substance Use and Misuse*, *45*(3), 437–450.

Molich, R., Chattratichart, J., Hinkle, V., Jensen, J., Kirakowski, J., Sauro, J.,…Traynor, B. (2010). Rent a car in just 0, 60, 240 or 1,217 seconds?—comparative usability measurement, CUE 8. *Journal of Usability Studies*, *6*(1), 8–24.

Muscanell, N. L., & Guadagno, R. E. (2012). Make new friends or keep the old: Gender and personality differences in social networking use. *Computers in Human Behavior*, *28*, 107–112.

Norman, K. L. (2008). *Cyberpsychology: An introduction to human-computer interaction*. New York, NY: Cambridge University Press.

Ofcom. (2016). Adults' media use and attitudes report 2016. Retrieved from https://www.ofcom.org.uk/__data/assets/pdf_file/0026/80828/2016-adults-media-use-and-attitudes.pdf

Oppenheim, A. N. (1992). *Questionnaire design: Interviewing and attitude measurement*. London: Pinter.

Prensky, M. (2001). Digital natives, digital immigrants. *On the Horizon*, *9*(5), 1–6.

Rafaeli, S., & Ariel, Y. (2007). Assessing interactivity in computer-mediated research. In A. N. Joinson, K. Y. A. McKenna, T. Postmes, & U. D. Reips (Eds.), *The Oxford handbook of Internet psychology* (pp. 71–88). New York: Oxford University Press.

Spector, P. E. (1992). *Summated rating scale construction: An introduction*. Sage University Paper Series on Quantitative Applications in the Social Sciences, No. 82. Newbury Park, CA: Sage.

Stahlberg, D., & Frey, D. (1996). Attitudes: structure, measurement and functions. In M. Hewstone, W. Stroebe, & G. M. Stephenson (Eds.), *Introduction to social psychology: A European perspective* (2nd ed.) (pp. 205–239). Malden, MA: Blackwell.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston, MA: Pearson.

Tsai, C.-C., Lin, S. S. J., & Tsai, M.-J. (2001). Developing an Internet attitude scale for high school students. *Computers and Education*, *37*(1), 41–51.

Van Deursen, A. J., van Dijk, J. A., & Peters, O. (2011). Rethinking Internet skills: The contribution of gender, age, education, Internet experience, and hours online to medium- and content-related Internet skills. *Poetics*, *39*, 125–144.

Weiser, E. B. (2000). Gender differences in Internet use patterns and Internet application preferences: A two-sample comparison. *Cyberpsychology and Behaviour*, *3*(2), 167–178.

Weiser, E. B. (2001). The functions of Internet use and their social and psychological consequences. *Cyberpsychology and Behavior*, *4*(6), 723–743.

Whitty, M., & Gavin, J. (2001). Age/Sex/Location: Uncovering the social cues in the development of online relationships. *Cyberpsychology and Behavior*, *4*(5), 623–630.

Wildermuth, S. M. (2004). The effects of stigmatizing discourse on the quality of on-line relationships. *Cyberpsychology and Behavior*, *7*(1), 73–84.

Zhang, Y. (2005). Age, gender, and Internet attitudes among employees in the business world. *Computers in Human Behavior*, *21*, 1–10.

Zhang, Y. (2007). Development and validation of an Internet use attitude scale. *Computers and Education*, *49*, 243–253.

36

Munchausen by Internet

Aideen Lawlor

Case Study

A weblogger's wife logs on to a blogging site to tell the members that her husband has committed suicide. She describes how much her husband loved the group, how they joked about the group being his "mistress" and the invaluable support and advice he had received down through the years. She recounts the numerous issues and difficult life circumstances he had faced. He was a former methadone addict. He was sexually abused by his uncle. His mother was a lifelong addict arrested for child endangerment and died of AIDS. He had a 13-year-old son from whom he had become estranged during his addiction but they were making progress towards rebuilding the relationship. He had just found out he had a 14-year-old daughter and was slowly building a relationship with her. His father and step mother were emotionally abusive. He suffered from crippling social anxiety. Recently, he felt partially responsible for the murder of a girl because he innocently provided a statement that led to the release of her boyfriend who subsequently murdered her. His wife talks about how happy they were together and their little idiosyncrasies. Saying "leprechaun, leprechaun" meaning "I love you." How she could guess his mood by asking what music he was listening to. The comic strip he wrote for her when he was in jail. The motivational and spiritual postcards he would send to her. She goes on to describe in detail the events leading up to his suicide. She went to her parents for the weekend. Her husband did not go as he had previously had a conversation with her father about God, science, and religion that had led to her father questioning his faith. This in turn did not please her mother, so he decided to stay home. The next morning, he was due to meet a friend to play disc golf; his friend heard the car engine was running inside the garage and found that her husband had died from fume inhalation. Initially she believed he had not planned the suicide in advance, even questioning if it was a murder made to look like suicide. However, a few days later she received one of the postcards he liked to send her. He explained that he was struggling with his methadone addiction and had always said he would die before he went back on it again.

The Wiley Handbook of Human Computer Interaction, Volume 2, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

Underneath this post there are 682 comments from members of the weblog describing how shocked and stunned they are by the news and offering their condolences, below are some examples:

> I've just now seen this and I am typing through my tears. I am very, very sorry for your loss. His comments and posts always made me look at life in a brighter light, even if they dealt with dark issues. Thanks for sharing your story. I'll miss his presence and will never forget his openness and kindness.

> I am so terribly sorry for your loss. I would say "may his memory be a blessing to you," but it is you who have blessed us with the memories you have shared. You and he and everyone who loved him will be in my thoughts.

> What you have shared here is profound, as was your husband's life. It feels like a true privilege that we've all had the chance to share in it. I want you to know that part of my heart that has been numb for a very long time was opened up and lit aflame by reading and rereading your and your husband's words all day today. I will not forget them.

Six days after the shock post that her husband had committed suicide, the moderators made the announcement that they had discovered that the whole thing had been a hoax. They had become suspicious as there was no obituary notice, there were no condolences on social media and there were references to a friend socializing with the weblogger a few days after the claimed suicide. When confronted, the weblogger admitted that it was a hoax. The response to the exposure was one of incredulity and anger, but some were philosophical about the whole experience stating that the outpouring of support towards the weblogger's wife only demonstrated the goodness of the community. Below are some examples of the comments:

> This is seriously fucked up. I feel so disgusted knowing that this happened. Someone tried to take advantage of the community and that's really upsetting. I don't know what to say.

> The takeaway I got from that thread is that this is an amazing community full of people who have mostly never met, but still have nothing but love and support for each other.

> I was momentarily angry, but I think I'm mostly sad, because, you know, there are a lot of people going through a lot of shit, and it's hard to reach out, and now some of them who do take that really hard step, I won't believe or I'll be skeptical about because I was deceived in this fashion. I'll be a slightly less generous and more cynical person.

Introduction

This case study of online pseudocide, involving the use of a sockpuppet (additional fake personae used for deception) helps to illustrate the stages of an online hoax, the detailed nature of the ruse, the genuine fondness members have towards the perpetrator before the hoax is revealed, and the mixed reactions they have when they realize they have been duped. It is a typical case of Munchausen by Internet (MbI), which is not an uncommon occurrence in online groups. MbI was first identified by Dr. Marc Feldman (2000), he described it as the misuse of Internet groups by offering false stories of personal illness and crisis to garner attention and nurturing. MbI is the online expression of what was commonly called Munchausen syndrome but is now diagnostically referred to as Factitious Disorder (FD). FD occurs when a person consciously feigns an illness and/or induces the symptoms of an illness to occupy the sick role (American Psychiatric Association, 2013). It is important to note that although the behavior is the same, FD and MbI must be distinguished from malingering in that there is no external incentive for the behavior, only internal (Feldman and Eisendrath, 1996). Malingering also occurs online, when a person maliciously portrays themselves or another as being sick for external incentives (money and gifts). The *DSM-5* also recognizes a second type of FD: FD imposed on another, where a person fabricates or produces symptoms in another person under their care. The online expression of FD imposed on another is MbI by proxy (McCulloch and Feldman, 2011). MbI by proxy typically involves pretending a relative or close friend is sick and posting about the experience online on their behalf, as a supporter and/or creating a sockpuppet representing the sick person. There is also a fourth type of online deceptive behavior that involves occupying the sick role, which was proposed by Pulman and Taylor (2012): MbI by trolls. This is a more sinister form where the aim is to disrupt and abuse the online group for enjoyment. The motivation may be similar to gaming community "griefers," who kill team mates and abuse more vulnerable players purely for enjoyments sake (Adrian, 2010).

As seen in the case study, these ruses are not just confined to trying the occupy the sick role, they also involve a lot of additional embellishment. This is referred to as pseudologia fantastica and with regards to MbI it typically involves false claims of victimization (Sadock & Sadock, 2011; Feldman & Birmingham, 2000). The use of pseudologia fantastica helps to create a captivating and compelling story that catches the reader's attention, which in turn feeds into the perpetrator's need for attention and sympathy. King and Ford (1988) describe the four characteristics of pseudologia fantastica: (a) the stories are not completely improbable and are frequently grounded in a matrix of truth; (b) the stories are enduring; (c) the stories are self-aggrandizing; (d) when confronted with evidence the perpetrator can acknowledge the falsehoods. Another important characteristic noted by Dupre (1909) is that while the themes of the stories may be varied, they will most certainly revolve around the hero/victim role. Therefore, MbI is not just limited to occupying the sick role but also can involve the occupation of the hero and/or victim role.

Motivation

The motivation for engaging in the four online deceptive behaviors (MbI, MbI by proxy, MbI by Trolls and malingering) are based on a very limited number of first-hand accounts from those who confess. Therefore, theories regarding motivation are mainly dependent on inferences from what is known about their offline equivalents. As will be discussed later, these inferences are problematic and ultimately independent research is required. However, the following will discuss what is currently known about each of the online deceptive behaviors based on extrapolations from FD and the limited first-hand accounts that are available.

Munchausen by Internet

Ayyer and De Sousa (2014) offer a rare insight into the perpetrators of MbI through the case study of an 18-year-old male who was brought for a psychiatric consultation by his parents. This was prompted after his deceptive behavior was accidentally exposed. A member of an online support group for cancer contacted the perpetrator's family trying to encourage him to go to the support group gathering in his town. It then emerged that he had been posting regularly on an online cancer support group where he falsely claimed to have terminal stomach cancer. He had even befriended one member who was also terminally ill and they had agreed to offer each other companionship. When confronted by his parents he was callous, mocking the members for their naivety and alluding to the incident as a mere prank. However, on further investigation it emerged that the motivation for the behavior was more deep rooted, as seen from the extract below:

> He felt that his parents did not care for him or understand him and he could not share his feelings with others. He believed ill people got a lot of care and attention from others and lying on internet sites made him feel wanted. He believed it was a harmless way of getting attention and fantasized about group members replacing his family. He had no insight that this was a psychological problem.

Similarly, Swains (2009) reports on a rare case described by Feldman where the perpetrator had also tried to explain her motivation:

> I have never felt more loved and cared for in my entire life. I suddenly craved for everyone's attention, love, care, concern and affection…It became very appealing to me. I decided to play with it more. I do not know how or why, I just did.

The motivation for MbI expressed in both cases was to assume the sick role for its psychological benefits. This is also speculated to be the primary motivation for MbI's offline equivalent, FD. Tasman and Mohr (2011) state that the need to occupy the sick role for these psychological benefits is because of predisposing factors that result in psychological deficits. The predisposing factors are thought to be rooted in childhood trauma/neglect, which in some instances can also result in the development of personality disorders. In particular, there is a strong comorbidity between FD and borderline personality disorder (Sadock & Sadock, 2008). There are a range of psychological deficits believed to be associated with FD but these are largely unsubstantiated (Ford, 1996). The psychological deficits most likely associated with enacting MbI include people with a poor sense of self using the sick role as a structure to develop their identity around, having unmet dependency needs met through the attention, caring, and nurturing associated with the sick role and, for those who have experienced trauma (childhood illness, abuse), creating a fictional trauma as a way of getting back control and mastering their real trauma (Fleming & Eisendrath, 2011).

Munchausen by Internet by proxy

The most common type of MbI by proxy reported in a study by Lawlor and Kirakowski (2017) involves a parent, typically a mother, pretending she has a child who is sick. They either blog about their experiences or join an online group and start posting about their tribulations. It is important to note that often online deception is not exclusive and can be accompanied with offline deception. There have been several cases where mothers were posting online about their sick children in addition to intentionally poisoning their children offline. The highest profile and intricate cases of MbI by proxy, which included a first-hand account from the perpetrator as to their motivation, was the Warrior Eli blog. The ruse ran for almost a decade. The fake blog was originally set up for the son of John and Emily Dirr, to chronicle his battle with childhood cancer. It was given credibility through the creation of 71 sockpuppets in the form of friends and relatives who communicated back and forth with the Dirrs. Suspicions were raised due to the series of tragedies that befell the family, culminating in Emily Dirr's death in a car crash on Mother's Day. She miraculously managed to stay alive long enough to give birth to the family's eleventh child. The hoax was exposed with the help of a blog set up by Wright (2013), when the pictures of the Dirrs and their various friends and relatives were found to be stolen from other people's social media accounts. Once confronted, the perpetrator admitted to the hoax and was revealed to be a 22-year-old female medical student training to be a doctor. She offered this apology via Wright's (2013) blog:

> To whom I have hurt: I am deeply sorry for all the pain I have caused everyone. It was never my intention to do so. This all started 11 years ago when I was a bored 11-year-old kid looking for an escape from the pain and heartache I saw in my own family. It started almost as a fiction writing, but the more time I spent escaping to it, the more "real" it became. I am so sorry it hurt so many real families, and so many people out there.

Apart from these rare first-hand accounts of the motivation behind MbI by proxy there is little known. Furthermore, the sincerity and accuracy of such confessions and apologies also must be met with skepticism. For example, in exposing the Eli Warrior hoax, Wright (2013) stated that the perpetrator in their initial confession had told them a sob story that turned out to be untrue. As there has been no independent research on MbI by proxy, theories regarding motivation are inferred from FD imposed on another. Those who perpetrate FD imposed on another do so to fill a psychological deficit, as is the case with FD. They seek to attract attention, sympathy, care, concern and a sense of importance from being the parent of a sick child as well as to manipulate and control others (Lasher & Sheridan, 2004).

Munchausen by Internet by trolls

Pulman and Taylor (2012) have argued that there is a need for a new category, MbI by trolls, where the motivation is primarily rooted in a comorbid personality disorder. It is purely for enjoyments sake to feed sadistic and / or narcissistic personality tendencies. Swains (2009) reports that Feldman also expresses the belief that there is a more sinister subset of MbI involving a preexisting personality disorder that primarily involves the manipulation of online groups and their members. Feldman argues that in some cases of MbI the feigned scenarios tend to be highly emotive and are followed by a superficial apology when exposed, suggesting an element of sadism. There is also a sense of being "in control" by controlling the thoughts and reactions of others online. Similarly, Feldman and Eisendrath (1996) note that FD is linked with comorbid personality disorders, in particular borderline personality disorder (Feldman & Eisendrath, 1996). The borderline personality is characterized by unstable interpersonal relationships, rage, impulsiveness, self-harm / suicide attempts, rapid changes in mood, and switching between idealizing and belittling others. When combined with FD, duping others allows them to focus their anger, thus temporarily stabilizing their mood. Duping online group users by falsely occupying the sick, victim and / or hero roles could be another outlet for those with a comorbid borderline personality disorder to focus their rage resulting in MbI by trolls.

Malingering

Malingering is distinct from the other types of deceptive online behavior involving the occupation of the sick, victim and / or hero role in that the motivation is purely for external incentives (Hamilton, Feldman, & Janata, 2009). Based on observations, the most common type of ruse seems to involve an element of deception by proxy. The online malingerer creates an emotive story by pretending their child is ill. They then extort money from members of the online group by claiming it is required for treatments, travel expenses, or gifts for the child. The distinction between malingering and FD is not always clear cut, however. This is highlighted by Worley, Feldman and Hamilton (2009), who argue that there is a difference between motivation and consequence. Therefore, a person with FD might find themselves in receipt of external incentives that accompany the sick role, but this was not their primary motivation. This too applies to MbI. Lawlor and Kirakowski (2017) reported on cases where online group members offered to and sent money and gifts, which were not solicited by the person with MbI or MbI by proxy. In fact, there were instances of the perpetrators encouraging the online group users to send money to charity instead.

Issues with applying extrapolations from FD to MbI

Although, as discussed, there are variants of online deceptive behavior involving the occupation of the sick, victim and / or hero role, for the purposes of the remainder of the chapter they will be referred to using the umbrella term MbI. At this point it is important to address the problem of applying extrapolations from FD to MbI. The most central problem is that there is a dearth of information available about FD given the elusiveness of sufferers, and what is known is based on inferences from observational case studies. These inferences have been directly called into question from a study by Lawlor and Kirakowski (2014). They sought to overcome the lack of first-hand accounts by analyzing the posts of people with FD discussing their experiences on an online support group. The findings contradicted the traditional theories of FD on several fronts. First, while the deception involved is intentional the motivation has always been assumed to be unconscious (Feldman & Eisendrath, 1996). However, Lawlor and Kirakowski (2014) found that sufferers self-reflected and speculated as to what their motivations were (affection, enjoyment, coping mechanism, sense of achievement, validation of genuine illness, identity issues, external gains, episodic triggers) indicating that motivation is not unconscious. Secondly, given the perception that motivation to assume the sick role is unconscious, it is also assumed that people suffering from FD do not experience negative symptoms, are unperturbed by their behavior and therefore do not feel the need to recover (Hamilton, Feldman, & Janata, 2009). Again however, the findings from Lawlor and Kirakowski (2014) contradict these two assumptions; sufferers described experiencing a wide range of negative symptoms (poor mental health, guilt and shame, addiction, dissociation from real self, negative self-perception, isolation, reduced quality of life and ostracized), and were eager to recover but employed their own self-recovery strategies rather than seeking professional help due to a range of fears around disclosure. These findings completely contradict traditional theories of FD, which by extension have also been tenuously applied to MbI. In addition, Bass and Halligan (2007) have argued that the fixation on diagnosing and distinguishing FD based on motivation is inherently problematic as motivation is highly variable and impossible to establish. Furthermore, Ford (1996) notes that there is a multitude of motivations put forward to explain FD and they are unsubstantiated. Instead Bass and Halligan (2007) believe the focus should be on the underlying psychiatric problems that precipitated the deceptive behavior. The question of motivation is further confounded by the tangled issue as to whether MbI and its variants are predominantly standalone disorders, deserving their own diagnostic subcategory in the DSM, or whether those with FD are simply using the Internet as another means of supplementing their offline disorder.

Given the litany of problems surrounding motivation and traditional theories of FD, it is important that research on first-hand experiences of MbI and its variants is conducted to establish if it is a standalone disorder and to determine its defining characteristics, motivation and beyond. However, until such time that research moves beyond case studies, our current understanding of MbI and its variants needs to be accompanied by the caveat that it is limited by extrapolations from traditional theories of FD, which are inherently problematic.

Effect of MbI on Online Groups and their Members

There have only been two studies examining the effect of MbI on online groups and their members. Feldman (2000) reported the effect that exposing hoaxes had in four case studies of MbI, and Lawlor and Kirakowski (2017) conducted a grounded theory analysis of 600 members of online support groups discussing their experiences and perceptions of MbI. Across the two studies, those who were duped experienced a variety of emotions. In Lawlor and Kirakowski's (2017) study, members reported feeling emotionally manipulated and violated. One member aptly described those with MbI as being "emotional vampires." Members described being emotionally invested in the MbI perpetrator, to the extent of genuinely caring for and emotionally supporting that person throughout various trials and tribulations. When deaths occurred as part of the ruse they genuinely grieved. The manipulation and violation was particularly felt by those who were going through the same issues being portrayed by the MbI sufferer and believed they had found an emotional ally to share in their difficult experiences. Once the hoax had been exposed they were left feeling humiliated, betrayed, and angry. Feldman (2000) also reported a similar reaction: those duped reportedly tried to contact the MbI sufferer to express their anger or sadness. They even tried to seek revenge by making contact through the sufferer's real-world contacts, such as employer or college, and they fantasized about or tried to devise a face-to-face confrontation. For those who had formed a close relationship with the MbI sufferer there was also the additional fear that personal information they had shared could be misused. In contrast, Feldman (2000) also reported that some were unperturbed by the experience and were in fact amused by the sophistication and audacity of the tales spun. Lawlor and Kirakowski (2017) also found that there was a subsection of online group members who viewed the behavior of MbI sufferers as harmless, found it amusing to watch the scenarios unfold, and could not understand why others were upset by the behavior. However, they were met with criticism from the majority who viewed the behavior as far from benign.

The presence or the threat of the possible presence of MbI sufferers also influences the online group. Feldman (2000) reported on how the exposure of a hoax can lead to division in the group between those who believe it was a hoax and those who disbelieve. Some members will remain in the group to process the variety of emotions they have experienced and others leave in disgust. In addition to the fallout when MbI hoaxes are exposed, Lawlor and Kirakowski (2017) also found that online group members were concerned about the threat posed by the possible presence of MbI in their groups. It created an air of suspicion and speculation; members described a nagging feeling that there were several undetected MbI perpetrators in their group and this discouraged them from participating fully. Those who had already been duped were particularly discouraged; they were hesitant to share their experiences, confide in others, or offer support in case of being duped. This loss of trust and air of skepticism led some members to believe that the sense of community was being eroded. This erosion of trust has a potentially negative effect for online groups, particularly online support groups. Klein and Dinger (2008) noted that, for members to benefit from participation, they must first make a risky emotional investment in the group. This leap of faith requires trust in other members without it they will not fully commit to the group or may decide not to participate at all.

Beyond the erosion of trust was the fear that suspicions would spill over into witch hunts. Lawlor and Kirakowski (2017) found that several members reported that they had been falsely accused of being fakes and were bullied. Considering these revelations, others began to question if they had ever been suspected of being a fake. In fact, members stated that they had censored, altered or restricted what they shared with the group if they felt it was consistent with the perceived criteria of being a fake—for example, dramatic, too many coincidences, atypical, and so forth. This fear of being falsely accused also led to members to ponder if genuine people were being discouraged from seeking help from the group. This finding creates an interesting paradox. On the one hand, online support groups have been heralded for their ability to facilitate honesty and self-disclosure through the disinhibition effect, thus expediting the therapeutic benefits of support groups (Barak, Boniel-Nissim, & Suler, 2008). However, on the other hand, Lawlor and Kirakowski's (2017) research suggests that if the perceived threat of MbI is high, leading to a fear of false accusations, this has the potential to inhibit honesty and self-disclosure, perhaps to the extent that others are afraid to participate at all. Under these conditions, anonymity leads members to become more closed rather than open. In addition, one of the corner stones of support groups is an atmosphere of nonjudgment and acceptance to allow for the unbridled sharing of difficult emotions (McCarthy, Kupiers, Hurry, Harper, & LeSage, 1989). If members are in fear of being falsely accused of being fake, their ability to share freely is impaired. It is important to note that Lawlor and Kirakowski (2017) found that the members discouraged witch hunts without solid evidence for these very reasons. They recognized the deleterious effect it has on the supportive environment fostered in online groups.

Although there is the potential for the perceived threat of MbI to discourage unbridled participation in online groups by eroding trust and creating an environment of suspicion and fear, Lawlor and Kirakowski (2017) found that some online group members were resilient against such threats. This underscores the importance of the online group to them. While these members offered sympathy towards those who had been duped, they felt that most members were genuine. They also felt that they would prefer to give the benefit of the doubt rather than ignore someone who needs support. Furthermore, they believed that even if the person they were supporting was an MbI perpetrator, someone else in a similar situation portrayed by the perpetrator could read their response and benefit from it.

There are other negative impacts of the presence of MbI perpetrators in online groups, which go beyond emotional manipulation and require further research. They potentially adversely affect recovery from the topic of physical or mental health issue discussed by the online group. One of the most serious consequences of MbI for those who are unwittingly exposed to their deceptions, is following the misinformation they spread. As part of the ruse, MbI perpetrators may provide false information about their personal history, medical advice they have been given, medical tests, treatments, and the progression of illness. They may also offer other members of the online group medical advice based on their "experiences." Mo and Coulson (2014) note that the experiences shared in online support groups have been shown to be used to inform health care and disease-management decisions. The inclusion of misinformation provided by MbI perpetrators could have negative health ramifications if used in the decision process. In addition, the "experiences" that MbI perpetrators share will also offer little hope to those going through similar experiences because the experiences will tend to be atypical, the condition will deteriorate to receive more sympathy and attention, and in extreme cases end in pseudocide. In contrast instilling hope through sharing positive stories of recovery is central to the success of support groups (Kurtz, 2015). So, for those who develop close relationships with MbI perpetrators, or if indeed the MbI perpetrator dominates discussions in the online group, the recovery of members could be negatively impacted through the preponderance of negative stories shared.

Detecting MbI

As MbI has a clear negative effect on online groups and their members, detecting the behavior early would be beneficial in helping to alleviate the distress it causes. In addition, it may improve the benefits that can be obtained from online groups by helping to limit the spread of misinformation and atypical negative stories. In cases of malingering, it would help to prevent extortion and fraud. Online group members are not the only people who would benefit. Research is increasingly being conducted into the efficacy of online support groups for mental and physical health issues and to date the findings have been mixed (Eysenbach, Powell, Englesakis, Rizo, & Stern, 2004; Griffiths, Calear, & Banfield, 2009). One limitation of such research is the unwitting inclusion of MbI sufferers who are unlikely to recover from their feigned issues; in fact they are likely to deteriorate. A method of detecting MbI sufferers so they could be removed during data collection would increase the validity of findings.

Research has been conducted into the features associated with MbI by Feldman (2000) and Lawlor and Kirkowski (2017). Feldman (2000) lists a series of clues that are associated with MbI and notes that detection of MbI is difficult as in many instances there is a mix of facts and lies used in the ruse:

- Consistently copies from others posts, text books, or health-related websites.
- The posting behavior (length, frequency, duration) is inconsistent with the severity of illness being claimed.

- The characteristics of the illness and treatment are like caricatures due to misconceptions.
- Oscillating from near to death to miraculous recovery.
- Fantastical personal claims, which are contradicted or disproved.
- Constant dramatic events, which escalate when attention towards them wanes.
- Complains that the other members are inattentive and this is compromising their health.
- Avoiding telephone contact by offering odd excuses.
- Feigned blitheness about a crisis, which will instantly attract attention.
- The use of sockpuppets—people posting on their behalf who happen to have the same unique writing style.

Lawlor and Kirakowski (2017) found that members referred to two types of cues: those that made them initially suspicious, and evidential cues that confirmed their suspicions. Cues that sparked suspicion included intuition, there was a feeling that something was "off" because the suspected MbI perpetrator could not emulate the emotions because they had not lived through the experience. As was also reported by Feldman (2000), Lawlor and Kirakowski (2017) found that stories that were dramatic and atypical also arose suspicion amongst members. Stories involving a constant stream of far-fetched tales, negative events, and containing contradictions and inconsistencies (e.g. the timeline of events is unrealistic, actions and behaviors don't match that of a person in their position, and posters trip themselves up trying to maintain lies). However, members also felt that these cues were applicable to those who are genuine and in fact they argued that these criteria could easily apply to their stories. Therefore, evidential cues were needed to confirm suspicions and members noted a variety that had been used in exposing hoaxes. These included stolen photos and identities where MbI perpetrators had stolen pictures belonging to other people's social media accounts, including pictures of their children, or where they had stolen the person's whole identity. This was particularly upsetting for those whose children's photos had been used to create a fake story about a sick child, which in some instances ended in the child's death. Another evidential cue was that the person did not exist offline—this was also noted by Feldman (2000). For example, there is no obituary; those who try to send gifts or visit the hospital are told there is no such patient, and if the death is dramatic there is no report of it in the news. The final evidential cue was shared IP addresses when sockpuppets were being used: multiple members would appear to be posting from the same IP address. Unfortunately, it takes a long time for these evidential cues to surface and often it requires extensive research from moderators and members to uncover it, based on their initial suspicions. In Lawlor and Kirakowski (2017) study online group members also discussed what would make them believe that a person was genuine. Some felt that meeting the person in real life or other members of the group vouching for the person would suffice. However, others argued that the members vouching for the person could be sockpuppets and there were cases of the person with MbI carrying the ruse over into their offline world.

Ultimately, detecting MbI based on observations is highly problematic. First, their stories are often a mix of truth and lies making it difficult to separate the two. Secondly, cues that raise suspicion or are believed to be indicative of genuineness are perceived by members to be equally applicable to those who are genuine and those who are fake. Thirdly, evidential cues often take time to uncover by which time the damage to the online group and its members has already been done. There is some promise however in looking towards more automated methods of online deception detection, by developing a classifier to identify MbI using feature selection and machine learning techniques (Lawlor & Kirakowsi, 2018). This method has already been applied successfully to computer-mediated communication using a variety of features to help aid in the detection of child exploiting chats, deceptive opinions in reviews, stylistic deception, gender deception, spam detection, and malicious profiles (Afroz, Brennan, & Greenstadt, 2012; Alowibdi, Buy, Yu, Ghani, & Mokbel, 2015; Fire, Katz, & Elovici, 2012; McCord, & Chuah, 2011; Miah, Yearwood, & Kulkarni, 2011; Ott, Choi, Cardie, & Hancock, 2011).

Managing MbI

Lawlor and Kirakowski (2017) found that members of online groups were very pessimistic about the ability to manage MbI successfully. They applauded moderators for the efforts they made in trying to protect the groups but ultimately they believed they had little control. Moderators cannot vet all members to guarantee their genuineness and they cannot monitor the group 24/7. Furthermore, even if they identify a case of MbI there are no sanctions they can apply to punish or discourage the behavior. One possible option suggested by members was that moderators would publicly out cases of MbI rather than just deleting the account or banning the perpetrator without any explanation. This is the current practice for most online groups. Members felt that exposing the perpetrator would help them to identify cues associated with the behavior, recognize repeat offenders, make those who had been duped aware of the deception, and shame the MbI perpetrator. In contrast several moderators and members shared their uneasiness with this approach. They believed that exposing MbI perpetrators would only feed their need for attention. Some cases of faking are also sensitive, for example involving minors and mentally ill people, where exposure would be inappropriate. Furthermore, the practice of exposing cases of MbI implies that the moderators are guaranteeing the genuineness of the remaining members, which is not the case.

Despite the reservations about publicly outing cases of MbI, there have been cases where the actual identity of the perpetrator has been exposed. This is one possible route to preventing and punishing the behavior. Feldman and Peychers (2007) reported on a case where a large online self-help website was abused by a member claiming to have experienced multiple life-threatening issues. Suspicions were raised due to the contradictory information he shared and his abrasive and threatening attitude once confronted. In response, a few members set up a website with the aim of exposing his behavior and identity to the public. The perpetrator took legal action to try and close the website because of defamation. The defense argued that the website was in the public interest and was completely truthful. The judge requested to see medical records to establish the veracity of the perpetrator's various illnesses. However, the perpetrator declined to provide the medical records and withdrew legal action. Members of online groups therefore have the power to expose deceptive behavior and the identities of perpetrators without fear of retribution, provided the claims are correct. This threat may act as a deterrent if members are made aware when applying to join the group. In addition, Pulman and Taylor (2012) suggest that MbI should be considered as a cybercrime rather than an accepted hazard of participating in online groups that cannot be controlled—particularly in cases where incorrect medical information is disseminated as part of the ruse, which subsequently results in the worsening of a health condition or fatality. In such instances perpetrators should be pursued through their IP addresses. In support, Feldman and Peychers (2007) argue that pursuing lawsuits is the ultimate way of turning the tables on those with FD; the same could be applied to MbI. Another novel solution to stem the medical misinformation associated with MbI was suggested by Witney, Hendricks, and Cope (2015). They argue that online support groups should employ a health professional who could provide evidence-based health information and correct any misinformation posted by members.

Until such time that MbI is treated as a cybercrime, members of online groups will continue to feel powerless. Lawlor and Kirakowski (2017) reported that, for now, members of online groups believe that the best policy is to report their suspicions to moderators and ignore the suspected perpetrator, thus allowing moderators to investigate and deal with the situation as they see fit. This is because in cases where the members had directly confronted the suspected perpetrator the results were mixed. The perpetrator either admitted to the behavior and apologized, left the group, or became abusive. This kind of mixed reaction after detection and confrontation was also reported by Feldman (2000):

- Vehemently protesting their innocence.
- Blaming the group e.g. they wouldn't have to lie if the group had been more supportive.
- Brusquely leaving the group and rejoining another to engage in the same deceptive behavior.
- Admitting to the deception but offering no apology or explanation.
- Admitting to the deception but mocking the members for their gullibility.

Additionally, Lawlor and Kirakowski (2017) found that members felt that because there was a risk of false accusations and the bullying of suspected perpetrators, which sometimes coincided with confrontation, it was best to leave it up to moderators to deal with the situation.

Future Research

Despite the significant negative impact that MbI perpetrators have on the functioning of online groups and the emotional distress it causes members, little research has been conducted into this online psychopathology. What is known is based on a handful of case studies and the remainder from problematic extrapolations from FD, which is poorly understood to begin with. It is therefore vital that future research goes beyond case studies and extrapolations, instead researching MbI directly in a more systematic way. The key questions that need to be addressed by such research include:

- What is the motivation for MbI and its variants?
- Is MbI deserving of its own subcategory in the DSM or is it another means of expressing FD?

- Should MbI be regarded primarily as a cybercrime or a mental disorder?
- How prevalent is MbI in online groups?
- What impact does MbI have on the dynamics and efficacy of online groups?

Given the potential for MbI to disrupt the therapeutic benefits of online support, it is also important for research to focus on how the design of online communities could be improved to minimize instances of MbI. For example, Lawlor and Kirakowski (2018) developed a text classifier capable of distinguishing text written by genuine user's and those with MbI, although they do stress that more research is required to compile a larger text corpus to train and test the classifier to a higher standard. However, given that the feasibility of a text classifier to detect MbI has been proven, a text classifier could be added to a forum as a plugin to alert moderators to suspicious users. Just as antispam plugins have been developed for online forums to detect both human and robot spammers using a combination of the links posted, IP and e-mail addresses and the post's content (Wynne, 2013). In addition, research could also investigate the possibility of improving the design of forums by including features that help to deter MbI altogether. For example, features have been put forward to deter trolls, including requiring registration and a probation period before posting, publicly displaying IP addresses and asking new users to post bonds that are refunded if they prove themselves to be reputable (Grohol, 2006; Hall, 2013; Kiesler, Kittur, Kraut, & Resnick, 2010). More novel methods could be developed to deter MbI based on a better understanding of this unique online psychopathology.

References

Adrian, A. (2010). *Law and order in virtual worlds: Exploring avatars, their ownership and rights.* New York, NY: Information Science Reference.

Afroz, S., Brennan, M., & Greenstadt, R. (2012). Detecting hoaxes, frauds, and deception in writing style online. *IEEE Symposium on Security and Privacy* (pp. 461–475). doi:10.1109/SP.2012.34

Alowibdi, J. S., Buy, U. A., Yu, P. S., Ghani, S., & Mokbel, M. (2015). Deception detection in twitter. *Social Network Analysis and Mining*, *5*, 1–16. doi: 10.1007/s13278-015-0273-1

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC: Author.

Ayyer, S., & De Sousa, A. (2014). Cyber-Munchausen syndrome: A case report. *Indian Journal for Mental Health*, *1*, 81–83.

Barak, A., Boniel-Nissim, M., & Suler, J. (2008). Fostering empowerment in online support groups. *Computers in Human Behavior*, *24*, 1867–1883.

Bass, C., & Halligan, P. W. (2007). Illness related deception: Social or psychiatric problem? *Journal of the Royal Society of Medicine*, *100*, 81–84.

Dupre, E. (1909). Mythomanic infantile, Un cas de fugue suivie de fabulation. *Encephale*, *11*, 117–126.

Eysenbach, G., Powell, J., Englesakis, M., Rizo, C., & Stern, A. (2004). Health related virtual communities and electronic support groups: systematic review of the effects of online peer to peer interactions. *British Medical Journal*, *15*, 1166–1170. doi:10.1136/bmj.328.7449.1166

Feldman, M. D. (2000). Munchausen by Internet: Detecting factitious illness and crisis on the internet. *Southern Medical Journal*, *93*, 669–672.

Feldman, M. D. & Birmingham, A. (2000). Munchausen by Internet. *Southern Medical Journal*, *93*, 1–3.

Feldman, M. D., & Eisendrath, S. J. (Eds.). (1996). *Somatoform and factitious disorders.* Washington, DC: American Psychiatric Publishing.

Feldman, M. D., & Peychers, M. E. (2007). Legal issues surrounding the exposure of "Munchausen by Internet". *Psychosomatics*, *48*, 451–452.

Fire, M., Katz, G., & Elovici, Y. (2012). Strangers intrusion detection-detecting spammers and fake profiles in social networks based on topology anomalies. *Human Journal*, *1*(1), 26–39.

Fleming, A. M., & Eisendrath, S. J. (2011). Factitious disorders. In A. Tasman, J. Kay, J. A. Liberman, M. B. First, & M. Maj (Eds.), *Psychiatry* (3rd ed.). Chichester: John Wiley & Sons, Ltd.

Ford, C. V. (1996). *Lies! Lies!! Lies!! The psychology of deceit.* Arlington: American Psychiatric Press.

Griffiths, K. M., Calear, A. L., & Banfield, M. (2009). Systematic review on Internet Support Groups (ISGs) and depression (1): Do ISGs reduce depressive symptoms? *Journal of Medical Internet Research*, *11*(3), e41. doi:10.2196/jmir.1270

Grohol, J. M. (2006). *Anonymity and online community: Identity matters.* Retrieved from: http://alistapart.com/article/identitymatters

Hall, J. M. (2013). *Simple solution to the troll/hacker/cyberbullying problem.* Retrieved from: http://getsatisfaction.com/imdb/topics/simple_solution_to_the_troll_hacker_cyberbullying_problem

Hamilton, J. C., Feldman, M. D., & Janata, J. W. (2009). The A, B, C's of factitious disorder: A response to Turner. *The Medscape Journal of Medicine*, *11*, 27.

Kiesler, S., Kittur, A. Kraut, R., Resnick, P. (2010). Regulating behaviour in online communities. In R. E. Kraut & P. Resnick (Eds.), *Evidence based social design: Mining the social sciences to build online communities* (pp. 125–179). Cambridge, MA: MIT Press.

King, B. H., & Ford, C. V. (1988). Pseudologia fantastica. *Acta Psychiatrica Scandinavica*, *77*, 1–6.

Klein, R., & Dinger, M. (2008). Trust in patient-centred e-health. In E. V. Vance (Ed.), *Patient centered e-health.* New York, NY: Medical Information Science Reference.

Kurtz, L. F. (2015). *Recovery groups: A guide to creating, leading, and working with groups for addictions and mental health conditions.* New York, NY: Oxford University Press.

Lasher, L. J., & Sheridan, M. S. (2004). *Munchausen by proxy: Identification, intervention and case management.* New York, NY: Routledge.

Lawlor, A., & Kirakowski, J. (2014). When the lie is the truth: grounded theory analysis of an online support group for factitious disorder. *Psychiatry Research*, *218*, 209–218. doi:10.1016/j.psychres.2014.03.034

Lawlor, A., & Kirakowski, J. (2017). Claiming someone else's pain: A grounded theory analysis of online community user's experiences of Munchausen by Internet. *Computers in Human Behavior*, *74*, 101–111.

Lawlor, A., & Kirakowski, J. (2018). Detecting Munchausen by Internet: Development of a text classifier using supervised machine learning. Manuscript Submitted for Publication.

McCarthy, M., Kupiers, L., Hurry, J., Harper, R., & LeSage, A. (1989). Counselling the relatives of the long-term adult mentally ill. I. Evaluation of the impact on relatives and patients. *British Journal of Psychiatry*, *154*, 768–775.

McCord, M., & Chuah, M. (2011). Spam detection on Twitter using traditional classifiers. In J. M. A. Calero, L. T. Yang, F. G. Marmol, L. J. G. Villalba, A. X. Li, & Y. Wang (Eds.), *Autonomic and trusted computing* (pp. 175–186). Berlin: Springer. doi:10.1007/978-3-642-23496-5_13

McCulloch, V., & Feldman, M. D. (2011). Munchausen by proxy by Internet. *Child abuse and neglect*, *35*, 965–966. doi:10.1016/j.chiabu.2011.05.020

Miah, W. R., Yearwood, J., & Kulkarni, S. (2011). Detection of child exploiting chats from a mixed chat dataset as a text classification task. In *Proceedings of Australasian Language Technology Association Workshop* (pp. 157–165). Canberra: Australasian Language Technology Association.

Mo, P. K., & Coulson, N. S. (2014). Are online support groups always beneficial? A qualitative exploration of the empowering and disempowering processes of participation within HIV/AIDS-related online support groups. *International Journal of Nursing Studies*, *51*, 983–993. doi:10.1016/j.ijnurstu.2013.11.006.

Ott, M., Choi, Y., Cardie, C., & Hancock, J. T. (2011). Finding deceptive opinion spam by any stretch of the imagination. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics* (pp. 309–319). Stroudsburg, PA: Association for Computational Linguistics.

Pulman, A., & Taylor, J. (2012). Munchausen by Internet: Current Research and Future Directions. *Journal of Medical Internet Research*, *14*(4), e115. doi: 10.2196/jmir.2011

Sadock, B. J., & Sadock, V. A. (2008). *Kaplan and Sadock's concise textbook of clinical psychiatry* (3rd ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Sadock, B. J., & Sadock, V. A. (2011). *Kaplan and Sadock's synopsis of psychiatry: Behavioural sciences/clinical psychiatry* (10th ed.). Philadelphia, PA: Lippincott Williams & Wilkins.

Swains, H. (2009, March 25). Q&A: Munchausen by Internet [Web log post]. Retrieved from: http://www.wired.co.uk/article/reports-of-my-death

Tasman, A., & Mohr, W. K. (2011). *Fundamentals of psychiatry*. Oxford: John Wiley & Sons, Ltd.

Witney, C., Hendricks, J., & Cope, V. (2015). Munchausen by Internet and nursing practice: An ethnonetnographic case study. *International Journal of Nursing and Clinical Practice*, *2*, 131–137. doi:10.15344/2394-4978/2015/131

Worley, C. B., Feldman, M. D., & Hamilton, J. C. (2009, October 30). The case of factitious disorder versus malingering. *Psychiatric Times*. Retrieved from http://www.psychiatrictimes.com/munchausen-syndrome/case-factitious-disorder-versus-malingering/page/0/1

Wright, T. (2013, January 10). Emily Dirr and Warrior Eli [Web log post]. Retrieved from http://www.warriorelihoax.com/2013/01/10/emily-dirr-and-warrior-eli/

Wynne, R. (2013). BbPress Complete: *A comprehensive guide to bbPress with clear and concise instructions on expanding your WordPress site with a community forum*. Birmingham, UK: Packt Publishing.

Part X
Communities

37

The Viability of Online Communities and Virtual Teams for Enterprise Clients

Tharon Howard

Introduction

In 2013, Towers Watson, a global professional services company, conducted a survey of 290 large companies in Asia, Europe, and North America and "found that 56% of the employers surveyed currently use various social media tools as part of their internal communication initiatives to build community—creating a sense that employees and leaders are in it together, and sharing both the challenges and rewards of work" (https://www.towerswatson.com/en-US/Press/2013/05/just-over-half-of-employers-using-social-media-tools-for-internal-communication). This survey suggests that debates from the early 1990s about whether or not electronic communities, virtual teams, and social networks are technologies that are "NSFW" or "not suitable for work" are nearly over. Gone are the days when it was standard operating procedure for companies to block network ports or Internet domains that their employees might use to "waste time" and network bandwidth participating in online communities. Naturally, most companies still have social media usage policies for their employees and they still protect their networks from some of the most damaging sites, which spread malware or botnets, like the infamous Aurora exploit that blew through most antivirus software in 2010 and infected half of the Fortune 100 companies (McMillan, 2010). Nevertheless, respected research firms like Gartner are predicting that "Enterprise social networks will become the primary communication channels for noticing, deciding or acting on information relevant to carrying out work" (Gartner says that 80% of social business efforts will not achieve intended benefits through 2015, 2013).

This chapter explores the growing trend among companies to create internal social media platforms, which their employees use to increase creativity and productivity. It explores three basic questions and contemporary research associated with each. First, why build virtual teams and online communities for internal, enterprise use? Second, how can organizations design their communities and networks for success? And last, how can communities be managed so that they last longer than 90 days and actually lead to successful outcomes for the companies and organizations that invest in them?

The Wiley Handbook of Human Computer Interaction, Volume 2, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

Why Build Virtual Teams and Online Communities?

The first and most obvious question is why so many major corporations and government organizations are building “enterprise social networks.” What factors are driving this interest? Chapter 3 of *Design to Thrive: Creating Online Communities and Social Networks that Last* (Howard, 2010, p. 29) offers the following 10 reasons why companies are building internal virtual teams and online communities to increase the productivity of their employees. He demonstrates that internal social networks will

- enhance and sustain an organization’s “intellectual capital”;
- increase creativity through crossfertilization;
- improve decision-making processes with “epistemic communities”;
- preserve institutional knowledge;
- flatten organizational hierarchies;
- improve retention and loyalty;
- reduce training and support costs;
- identify customer needs and new product opportunities;
- reduce travel costs and address problems “just in time”;
- provide a higher quality relationship with your organization.

What is particularly noteworthy about these 10 reasons is that the first five all have to do with the impact that online communities can potentially have on decision-making processes and knowledge production within an organization, while the last five have to do with cost savings resulting from that impact. For example, in the Unix Operations team at Blue Cross and Blue Shield (BCBS) in South Carolina, they are developing a new wiki for the organization in order to capture knowledge about processes system administrators to maintain and update their servers and databases. One of the BCBS’s problems, like so many organizations, is that when employees transfer to another business unit or leave the organization, all of the institutional knowledge they have accrued goes with them; all of their understanding of, say, security groups and operational processes on a particular Unix server goes with them. This loss of intellectual capital translates directly in to lost economic capital in terms of the time it takes to train a new sysadmin for those systems, not to mention the potential disruption of mission critical servers and databases that are lost while the new sysadmin is getting up to speed. The wiki preserves that institutional knowledge in the form of a keyword searchable social media forum.

What’s more, beyond merely protecting the institutional knowledge, BCBS also reduces its training and support costs through the wiki. New sysadmins don’t have to wait for training sessions to be offered before they can gain access to procedural information needed to do their jobs, nor do they need to travel to remote sites for training—all that sort of information is stored on the wiki where it can be accessed “just in time.” Blue Cross and Blue Shield also benefits by lowering its support costs for clients. In this particular case, the “customers” are primarily employees from other national and international business units within the organization who may also access the wiki directly to obtain answers to questions they may have about services available on the system. In other words, rather than having to contact a sysadmin by phone and incurring the costs of distracting the sysadmin from other important work (not to mention annoying the sysadmin with a frequently asked question), clients can search the wiki themselves 24/7 and get immediate answers.

Beyond these direct benefits and cost savings, the BCBS Unix Ops wiki also generates indirect, yet critical, advantages for the community that contributes to and supports it. Two of the principal benefits are enhanced creativity and improved decision-making processes, which result from the flattening of organizational hierarchies produced by online communities. In his research on creative problem solving in business, Ronald Burt (2000, p. 6) found that creative solutions come from individuals who use what he called "social capital" from different communities. He observed that:

> Information can be expected to spread across people in a market, but it will circulate within groups before it circulates between groups. A generic research finding is that information circulates more with than between groups—within a work group more than between groups, within a division more than between divisions, within an industry more than between industries.

Put another way, Burt is simply expressing the old problem that employees in an organization can get "siloed in" at their organization and become "cubical slaves" when they are denied access to the social capital available in online communities where information is shared between work groups and multiple business units. Burt found that creativity, on the other hand, emerges when groups are able to share social capital and, in so doing, learn how to solve business problems by applying techniques from one field to problems found in another. In a famous example of this, Gutenburg was able to invent the printing press by applying the techniques he had learned as a goldsmith to the problem of creating the hot metal type that allowed him to apply ink to paper. Similarly, in a more modern example, UX designers have borrowed the idea of creating customer journey maps from the fields of systems engineering and marketing in order to develop the increasingly popular User Experience Journey mapping technique. Companies can thus enhance the creative potential of their employees by creating online communities that allow their employees to obtain social capital from others—the Unix Ops team at BCBS could, for example, learn to apply techniques used by their peers in the Windows Ops unit to find creative solutions to their own problems.

How to Set Up Teams for Success?

Organizations will not experience any of the 10 benefits listed above, however, if they fail to set up the appropriate architectures for their teams, and sadly, this failure often seems to be the norm. In their 2013 survey, Towers Watson found that, of the 290 large and midsized companies that were using social media, "only four in 10 (40%) rated the use of social media technology as cost effective." Or to put this another way, just because you build a social media platform for your employees, it doesn't mean that "they will come" or, more importantly, it doesn't mean that they will use it in order to complete the kinds of activities you expected.

One of the principal reasons for this inefficiency is that the architects of social media platforms fail to recognize the importance of distinguishing between "virtual teams," "communities," and "social networks." Instead, because the tremendous commercial success of Facebook, Twitter, and LinkedIn, companies are trying to emulate these social networks with their internal social media platforms. However, as

Mark Granovetter has pointed out in his seminal and oft cited work on social networking theory, “The Strength of Weak Ties”:

> Most network models deal implicitly with strong ties, thus confining their applicability to small, well-defined groups. Emphasis on weak ties lends itself to discussion of relations *between* groups and to analysis of segments of social structure not easily defined in terms of primary groups (Granovetter, 1973, p. 1360).

In order to understand the difference between online communities and social networks, it’s necessary to use networking theorists’ distinction between “strong,” “weak,” and “absent ties” in a networking system. Basically, strong ties between individuals or what network theorists call “nodes” in a network are those nodes where the individuals are connected because they share a close, personal relationship. David Krackhardt (1992) argues that strong ties in a system have three characteristics:

- the individuals must have significant *interaction* with each other such as seeing each other frequently;
- the interaction must have occurred over a significant period of *time*;
- and there needs to be *affection* between the individuals in the nodes.

By way of contrast, “weak ties” are links between nodes where there is a lack of emotional connection and significant interaction. Put in somewhat subjective terms, an example of a strong tie would be individuals who are friends or colleagues who see each other every day at work. Weak ties, on the other hand, are acquaintances rather than friends; they are people who might work at the same company and might even see one another in the hallway, but don’t know each other’s names.

Social networks are built almost exclusively around strong ties. Users begin building their social network by locating and then “friending” family members and close friends. Individuals use, in sum, those strong ties that they can find in the networking system in order to create their own, personal social network. In a *Time* interview on July 2007, Mark Zuckerberg describes Facebook’s use of strong ties this way:

> Our whole theory is that people have real connections in the world. People communicate most naturally and effectively with their friends and the people around them. What we figured is that if we could model what those connections were, [we could] provide that information to a set of applications through which people want to share information, photos or videos or events. But that only works if those relationships are real. That’s a really big difference between Facebook and a lot of other sites. (Locke, 2007)

An important consequence of this, which is often overlooked, is that each “social network” on Facebook or LinkedIn is unique to the individual around whom it is created. Although many people refer to applications like Facebook as “*a* social network” and while companies may say that they are building “*an* internal social network” for their employees, the singularity of their terms is incorrect and misleading. Corporate social networking tools like Chatter, Socialcast, tibbr, or Yammer allow for social networking by individuals in the system but the actual social networks that are created by these systems have the strong ties of an individual at their core. The company doesn’t have a social network; rather, each employee has his or her own unique network on the corporate system. The downside of this observation is that corporate social networks can fail to produce many of the 10 benefits that were discussed in the previous section. For example, in terms of giving employees access to the kinds of social capital necessary for creative problem solving that Burt observed, the strong ties in social networks often don't prevent individuals from becoming "silo'ed in." As human beings, we tend to "friend" people who travel in the same social circles that we do, thereby creating a strong sense of homogeneity in our own, unique social network, limiting our access to interactions with people from other groups. Granovetter (1973) thus observed that systems based on strong ties tend to encourage interactions within groups rather than *between* them.

Unlike social networks, "online communities" (and their smaller counterparts "virtual teams") are characterized by their ability to create networks among nodes that have weak ties. Systems like Mediawiki, Listserv email distribution lists, Jive, PHP-Nuke, Drupal, Forumbee, and other online community tools don't require that members know each other personally. Indeed, it's quite common that participants in an online community don't even know the names of most of the others in the community. This isn't to say that there cannot be strong ties among some members of online communities and virtual teams; in practice, there usually are. However, the organizational structure of online communities isn't contingent on strong ties and depends instead on the mutual interests, shared goals, or common experiences of the members of the community. Because these weak ties allow for overlap with individuals who may come from different circles, disciplinary traditions, and groups than those reflected in an individual's social network, communities encourage interaction between groups in ways that social networks do not. This opportunity to interact with and obtain social capital from groups outside your own is the real strength of weak ties.

Now at this point, it may appear that there's little to be gained by building social networks in a company when the weak ties in online communities lend themselves so clearly to benefits like creative problem solving. However, such a conclusion would be an overgeneralization. After all, Facebook, Weibo, and Qzone wouldn't be the behemoths they have become if they weren't good for something and didn't fulfill some useful purposes. The actual point that should be observed here is that the architects and managers of enterprise level social networking systems need to be clear about the types of goals that they wish to achieve with their systems. In his book *Here Comes Everybody: The Power of Organizing Without Organizations*, Clay Shirky (2008) describes three levels of organizational complexity (sharing, cooperation, and collective action), which architects and managers can use to decide whether social network tools or whether online community tools will accomplish an organization's goals.

Sharing requires the least organizational complexity in Shirky's hierarchy, which is what makes it ideal for social networks. Because sharing doesn't necessarily require interaction between groups, the strong ties within a social network make it easy for members to spread pictures, ideas, or pretty much any type of information throughout the nodes in a social network. Indeed, this is what makes Twitter such a powerful tool for journalists. It also makes it a useful tool in corporate environments where sharing information is desirable. Consider, for example, the following anecdote from Charlene Li's (2015) somewhat ironically titled *Harvard Business Review* article "Why No One Uses the Corporate Social Network":

> At Red Robin, a chain of over 450 casual restaurants, Chris Laping, the CIO and senior vice president of business transformation, spearheaded the companywide implementation of Yammer, an enterprise social network. When the chain launched its Pig Out Burger in 2012, employees posted that the new menu item was getting panned by customers.

> Reviews flooded in and were funneled to executives and to the test kitchens at headquarters. "Managers started talking on Yammer about ways to tweak the Pig Out recipe and four weeks later we had an improved, kitchen-tested version to roll out to customers," Laping shared. "That's a process that would have taken 12 to 18 months before" (Li, 2015).

In this example, no real complexity in the organizational structure was required other than the strong ties between individuals. Information about the new menu item was shared and reshared through the flat, weblike organizational structure of individual social networks until eventually it reached managers who were able to act on the information.

Activities involving *sharing* don't require complex structures; a simple weblike organizational structure will suffice. However, when managers and architects decide that their companies require *cooperation* or *collective action*, then the organizational complexity needed requires weak ties. In cooperation, members of the group need to assume responsibility for different roles. Say, for example, a group has to cooperate in order to pull together a conference such as a World Usability Day event, which will be hosted by a local UXPA chapter at a nearby university. If all that was needed in order to put together this event was sharing information about when and where the event would be held, then social networks would be the best tool to use. However, the organizational complexity required for collaboratively producing even such a small event is much more multifaceted and intricate than the strong ties in a social network can maintain. A simple web of connections between nodes isn't sufficient. There must be coordination of roles and responsibilities that members in the organization will play. Someone has to locate and schedule a host facility for the event, someone has to arrange speakers and presenters for the event, someone has to develop a schedule, someone has to arrange to provide refreshments during breaks, someone has to assume responsibility for advertising the event, and so forth. Leadership is required in each of these areas, and in fact it is likely that there is a layer of coordination above the leadership of the teams responsible for speakers, publicity, and facilities. There is, in other words, a hierarchical structure where conference managers sit at the top coordinating the activities of the team leaders who are, in turn, responsible for the individual volunteers who provide the effort needed to line up speakers, create publicity, and reserve host facilities and presentation equipment. Collective action requires even more complex organizational structures because (unlike coordination) we're not merely dealing with individuals assuming responsibility and playing roles. Collective action involves unions, trade associations, governments, and professional organizations making agreements with other organizations or creating policies, codes of behaviors, bodies of knowledge, or certification practices for their memberships. As such, collective action goes beyond coordination because it can require that individuals sacrifice their sense of identity and make compromises for the greater good.

Shirky's three levels of organizational complexity needed for producing sharing, coordination, and collective action activities—combined with an understanding of Granovetter's weak and strong ties networking theory—provides managers and architects with a useful heuristic for helping decide whether to use social networks or online communities as part of their enterprise social media strategies. In *Design to Thrive* (Howard, 2010), the heuristic is summarized in Table 37.1. Using this

Table 37.1 Comparison of social networks and online communities.

A social network	*An online community*
Has an organizational structure focused around an individual user's one-to-one relationships.	Has an organizational structure focused around a shared purpose rather than one-to-one relationships.
Has weak secondary connections between members.	Has strong, predictable secondary relationships between members.
Allows its users to be members of many communities in the network at the same time.	Is distinct from other communities because of differences in purpose, policies, and computing environment.
Is good for sharing activities.	Is good for activities requiring sharing and cooperation.
Is less effective at activities requiring cooperation and collective action.	Is effective at providing the framework for activities requiring collective action.
Makes it easier for users to build communities.	Should not be confused with "adhocracies," "discussion groups," "forums," or "lists."

Source: Howard (2010, p. 11).

heuristic, managers and architects can avoid the poor decision making processes, which, as the Tower-Watson study showed ("Just over half of employers using social media tools for internal communication, Towers Watson Survey Finds," 2013), leads so few companies to experience a positive return on their investments in social media technologies.

How to Manage Communities for Success?

In her often-cited book *Online Communities: Designing Usability, Supporting Sociability*, Jenny Preece (2000) states that "Sociability is concerned with planning and developing social policies which are understandable and acceptable to members, to support the community's purpose" (p. 26). In other words, managing a community or social network involves creating policies that meet the members' needs; however, doing so requires an understanding of those needs and recognizing that not all members of a community are the same. Successful managers, therefore, need a thorough understanding of different membership models and theories. Consequently, this section will review: (a) management models based on membership life cycles, (b) models based on defining members by interaction types, (c) models based on managing members' stages of team development, and (d) management based on providing characteristics of successful online community experiences.

Members' needs can evolve and change over time, and so one of the most popular methods for considering the different needs of members is to build a membership life-cycle model. Perhaps the most popular approach to building a life-cycle model of online communities combines the online work of Amy Jo Kim (2000) with that of Etienne Wenger's (1998) influential studies of more traditional communities of practice and the "learning trajectories" individuals take through the communities. In her classic book, *Community Building on the Web*, Kim uses a life-cycle metaphor to generate her classification of membership types. Kim (2000, p. 118) describes five "successive stages of community involvement," and she illustrates below how each stage correlates with a different membership type:

1 Visitors: people without a persistent identity in the community.
2 Novices: new members who need to learn the ropes and be introduced into community life.
3 Regulars: established members [who] are comfortably participating in community life.
4 Leaders: volunteers, contractors, and staff who keep the community running.
5 Elders: long-time regulars and leaders who share their knowledge and pass along the culture.

In terms of Preece's "sociability" and setting policies for a community, the last two types of users in Kim's list are of particular importance. Elders, it should be noted, set the tone for discussions and can serve as monitors or informal censors of inappropriate behaviors. Leaders in a community are also clearly critical to the success of a community; hence Kim (2000, p. 145) identifies seven possible leadership roles that can be assigned, depending on the type of community that is operating:

- Support providers answer questions, help members solve problems they're having with the system.
- Hosts keep the key community activities (games, conversations, shopping, etc.) running smoothly.
- Greeters welcome newcomers, show them around, and teach them the ropes.
- Cops remove disruptive members and/or inappropriate content.
- Event coordinators plan, coordinate, and run one-time and regular events.
- Teachers train community leaders, offer classes, or provide tutoring.
- Merchants run shops, provide services, and fuel the community economy.

In 1992, Etienne Wenger began his work on communities of practice by conducting ethnographic studies of how people are socialized into communities such as quartermasters on U.S. navy ships, midwives in Yucatec, clothing tailors in Gola, and other professional communities. In his 1998 book *Communities of Practice*, he began to investigate how a novitiate's sense of identity might evolve over time as individuals move from entry-level beginners to accepted and established members of the community. Wenger (1998, pp. 154–155) called these shifting senses of identity "learning trajectories" and he identified five of them:

- *Peripheral trajectories.* By choice or by necessity, some trajectories never lead to full participation. Yet they may well provide a kind of access to a community and its practice that becomes significant enough to contribute to one's identity.
- *Inbound trajectories.* Newcomers are joining the community with the prospect of becoming full participants in its practice. Their identities are invested in their future participation, even though their present participation may be peripheral.
- *Insider trajectories.* The formation of an identity does not end with full membership. The evolution of the practice continues—new events, new demands, new inventions, and new generations all create occasions for renegotiating one's identity.

- *Boundary trajectories.* Some trajectories find their value in spanning boundaries and linking communities of practice. Sustaining an identity across boundaries is one of the most delicate challenges of this kind of brokering work.
- *Outbound trajectories.* Some trajectories lead out of a community, as when children grow up. What matters then is how a form of participation enables what comes next.

Because both Kim and Wenger take a life-cycle approach to the stages that members of a community can follow, and because both provide five stages, it has become popular to combine the two approaches. The result is a five-stage theory of membership:

1 Peripheral member (lurker)—These are those who visit the community but do not participate.
2 Inbound (novice)—These are newcomers learning the ropes of the community.
3 Insiders (regulars)—These are consistent members of the community who regularly post messages and interact with others.
4 Boundary (leaders)—These are the veteran and most influential members of the community; they are often the initiators of discussions and they set the tone and behavioral protocols for conversations.
5 Outbound (elders)—These are members who are gradually leaving the community—sometimes because they don't appreciate the direction the community has gone or sometimes due to alternative social commitments.

Obviously, this five-stage model privileges Wenger's approach over Kim's and does some violence to Kim's definitions of "visitors" (who are turned into lurkers and do not contribute to the community) and "elders" (who are said to be retiring from the community rather than sustaining it by helping to manage the tone of conversations). Similarly, Wenger's observations about how members on the "boundary" of communities can enhance the creativity and range of conversations in the community by establishing connections with other communities is lost in the combination of Kim's and Wenger's approaches. Nonetheless, the combination of the two approaches does yield a manageable and easy to visualize model of the process members go through over the course of their experiences with an online community. Dion Hinchcliffe (2008) on his ZD Net blogsite provides a compelling view of the combined approaches and the five stages in Figure 37.1.

In addition to life-cycle models, which look at the types of members and their needs from a diachronic perspective, there are also taxonomies of membership types, which classify members by the types of contributions they make to communities. Forrester researchers Charlene Li and Josh Bernoff have developed what they call the "social technographics ladder" in their book *Groundswell* (Li & Bernoff, 2008), which identifies types of community members by their contributions. There are six types or "rungs" on Li and Bernoff's ladder: (a) Creators, (b) Critics, (c) Collectors, (d) Joiners, (e) Spectators, and (f) Inactives. Table 37.2 from *Design to Thrive* (Howard, 2010), explains the characteristics for each member type and then expands on Li and Bernoff's work by illustrating the needs that community managers should meet in order to satisfy those members.

Understanding the needs of different membership types in an online community and creating policies that are consistent with their needs is one important approach to

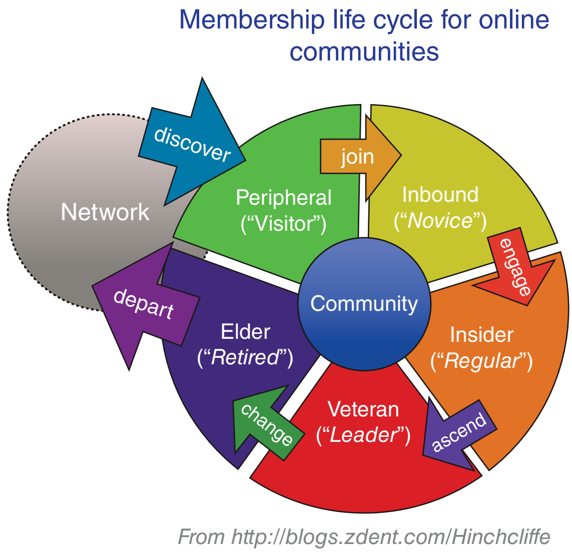


Figure 37.1 Visual from Hinchcliffe's blog.

Table 37.2 The social technographics ladder member types, characteristics, and influence needs.

Member type	*Characteristics*	*Influence needs*
Creators	Create the primary, initial content for a community or network Want to shape and persuade other members.	Need to know they're being read and have some impact
Critics	Critics will comment on the entry Creators post. Consensus builders	Need feedback, which helps them determine how the values in the network or community are lining up.
Collectors	Add value to social networks and communities because they sort the content created by Creators and Critics into ordered and ranked categories	Need to have questions sent to them asking if they have information in their archive that will help the community. Need to hear that their social bookmarks or archives or collecting behaviors are valued by the community
Joiners	Have influence simply by virtue of the profiles that they choose to fill out and complete	Need to see that other people have viewed their profile information. Need to see that their mere presence had influence.
Spectators	This is the group that benefits from and "consumes what the rest produce" (Li and Bernoff, 2008, p. 45) Implies appreciation for the content generated by the Creators and Critics	Need to see that Creators and Critics appreciate the time that they invest reading their blogs or postings or watching their videos.
Inactives	Expend no energy on behalf of the social network and don't participate	Have all of the influence needs of the previous five membership types—but these needs aren't being met.

Source: Howard (2010, p. 90).

managing communities for success. Creating policies and practices that, for example, celebrate the contributions of "elders" in a community can dramatically enhance the sustainability of a community. However, it's also possible to apply management techniques for face-to-face teams to online communities and virtual teams. Unfortunately, according to Hill and Gruner (1973), there are over 100 theories and techniques for managing small groups and teams that could be applied to virtual teams—far more than can be discussed in this chapter. Yet, one classic example of this type of application to the online world is the use of Bruce Tuchman's four stages of team development. Tuchman, who was a submariner and therefore interested in team building for military purposes, conducted a meta-analysis of 50 empirical studies on team building and collaboration. He found that in all the studies he surveyed, successful teams went through four distinct and inevitable phases of team development. These famous four phases are: Forming, Storming, Norming and, Performing (Tuckman, 1965).

During the "forming" phase of team development, members' interactions with one another are primarily governed by social protocols and cultural practices. Team members are trying to get to know one another and, as a result, engage in "polite" conversations intended to avoid conflict and emotionally charged discursive practices. Even though individual members may be annoyed with or may disagree with assertions made by other members of the team, during the forming phase they withhold comments and judgmental behaviors. After the initial forming stage, however, teams enter into the "storming" phase.

Storming doesn't necessarily mean that members participate in shouting matches or that they attempt to beat down one another, but it does signify that team members are engaged in an effort to first identify conflicts and sources of disagreement that may exist in the team and second to work through those conflicts. Some members, for example, may believe that a single group leader will make the final decisions regarding issues and problems that the team may face after first having obtained input from each member of the team. Others may assume that the team is using a consensus model for its decision-making process. Indeed, there can often be significant confusion regarding what it means to arrive at consensus on a team. Some, for example, may believe that consensus is unanimous agreement; others assume that consensus is the majority perspective; while still others may believe that "consensus forms in the mind of the chair." These differences in the assumptions that govern team members' behaviors and expectations of how their peers can and should contribute to the team's efforts, need to be identified and articulated explicitly. Tuchman found that teams that failed to "storm" and fail to articulate the assumptions that govern their expectations are teams that typically underperform. Storming is, thus, a critical phase in the development process, and for managers it is often a good idea to review these four phases with teams so that they are not intimidated by the "storming" phase.

Once teams have identified problems and assumptions through the storming phase, they can then enter into the "norming" phase. As the term suggests, in the "norming" phase teams will decide on the social norms that they will use to make decisions and set expectations for one another, create roles for members to play, and so forth. Then, once the norms have been set and are understood by each member of the team, the teams can then enter into the "performing" phase where they cooperate with one another in order to achieve their shared goals. Management systems like Tuchman's four phases of team development can work well if the goal is to manage small, virtual teams that have relatively short time frames for completing cooperative or collective agreement activities.

However, for online communities, which don't have a clear end date and which can be expected to last for years, then team management tools like Tuchman's aren't viable. Unlike teams, the management of long-term online communities is more of an art than a matter of simple techniques. Still, one heuristic tool that community managers can use in order to help them make decisions about policies that they need to create and enforce is the RIBS theory found in *Design to Thrive*. RIBS is an acronym that stands for:

- Remuneration
- Influence
- Belonging
- Significance.

Each of these four terms represents an essential characteristic of successful and sustainable online communities. For each characteristic, there are at least 12 to 15 different strategies that community managers can utilize. Yet, the focus of the heuristic is on understanding each of the four elements in the RIBS heuristic since each represents a factor in successful and sustainable online communities that managers should seek to address.

The first concept, *remuneration*, deals with the benefits that members enjoy as a result of their participation in the community. At its most basic level, remuneration is the idea that people won't stay members in a community very long if they don't perceive benefits for doing so. These benefits don't need to be and usually aren't monetary. Instead, they are experiential and phenomenological. Although the term "remuneration" may suggest monetary rewards, the concept is actually more about experience and emotion—i.e., "scratching a social itch" we share as human beings. In the concept of remuneration community managers are encouraged to view communities as "decision-making engines" rather than special interest groups, hobbyist organizations, affinity groups, or even professional development forums. Remuneration theorizes that members of communities are rewarded for their participation in a community by social meaning making. We don't, for example, typically go to movies by ourselves. Instead, we go in a group and, in discussing the movie with other members of our trusted social circle, arrive at conclusions about the meaning and value of the movie. It's in this same sense that remuneration in online communities is experiential and phenomenological. In Western societies in particular, we like to believe that we, as individuals, make sense of the world around us when, in reality, most of our values, goals, and expectations of phenomena in the world around us are socially constructed. Thus, the remuneration principle asserts that "The key to long-term success is remembering that the most important remuneration you [the community manager] have to offer is the *experience* of socially constructing meaning about the topics and events your users want to understand" (Howard, 2010, p. 57).

The second RIBS characteristic of successful online communities is *influence*. Like remuneration, influence is once again an experiential concern, only here the issue is that members won't stay members long in a community if in their experience with a community they don't perceive that they have some influence both *within* and *over* the environment. In terms of the experience of influence within a community, it's usually enough to ask social media users how they feel when they post a message to an email distribution list or to a forum and no one responds. What is the emotional impact of posting say a vacation photo Facebook and not receiving any comments or likes? That experience of not having any influence with others in the social environment is devastating, and it's not one that members will tolerate for long. Because of experiences like these, the individual doesn't perceive that she has influence within the community. However, influence isn't only an issue *within* a community; it's also an issue when members don't perceive they have any control of what happens *over* the environment. Members' concerns with having influence over a community typically manifests itself in policies that govern privacy, intellectual property, shutting down trolls, and so on. For example, community managers may have set policies regarding the ways that members' messages can or should be used. And those policies may be inconsistent with the members' goals. The members' ability to influence and change those policies is what is at stake here. For example, when it was discovered that Facebook's policies regarding how the company could use users' images for advertising without members' consent, there was a real outcry from Facebook users. Similar concerns emerge when employees on enterprise systems discover that HR managers are monitoring their postings in order to assess morale and potential disciplinary problems. When employees discover this was going on, they stop posting in fear of retribution and concern that their posts could be used out of context. The influence heuristic shows that companies need to publicly announce the change in policies and modify their practices in order to retain employees in the community and to keep their participation active. Being responsive to users' needs and giving the perception that they have control within and over community is one of the criteria necessary for successful long-term communities.

The third RIBS principle is *belonging*. Like the previous two principles, belonging is an experiential concern because it is a heuristic that primarily addresses protocols and rituals that community managers develop in order to create a sense of belonging in the community for their members. One popular example of this in many current social media sites is the use of badges or ranks or "karma points," which show a member's level of engagement in a community. Members who have earned, for example, a green belt or a brown belt in a community have a sense of belonging and are encouraged to continue their participation. Initiation rituals are another practice that can help members achieve a sense of belonging. Having a formalized ceremony that welcomes someone who has completed a rite of passage which all the other members of the community have also shared, creates bonds of belonging and strengthens an individual's commitment to the community. Other methods of creating a sense of belonging involve the use of protocols that are unique to a community. Protocols are like secret handshakes or the display of symbol systems, which are ways to show your membership in a community. For example, in the online community UTEST, which has been one of the premier online organizations for user experience researchers and practitioners for nearly 25 years, a number of protocols have evolved. One of these are "casual Friday postings." UTEST has a very positive signal to noise ratio, and jokes or humorous anecdotes are assiduously avoided during a typical discussion of professional topics. In fact, members who fail to comply with this protocol have been censured for being a "noisy neighbor." But on Fridays members can ease off a bit and let their hair down. They post links to, for example, Dilbert cartoons that have UX issues or post funny anecdotes about their workplace experiences. As a result, starting an e-mail message on UTEST with "since it's Friday" functions like a secret handshake that indicates the person posting the message understands and shares a bond with other members of the community. In other words, the use of the Friday protocol creates a sense of belonging among both the original poster and readers of the message. In sum, community managers who use protocols, symbols and rituals to help their members experience a sense of belonging are far more likely to insure that the communities last and enjoy sustained success.

The final heuristic principle from RIBS is *significance*. This deals with the fact that members won't stay long in communities that they perceive as lacking in *gravitas* or significance. If participants sense that the community is trivial and unimportant then they're more likely to engage in disruptive behavior, which harms the community because it encourages other members to leave as well as the individuals themselves. As with the previous three heuristic principles from *Design to Thrive* (Howard, 2010), significance involves the lived experience of a community member. In order to help members experience "the significance" of a community, community managers can use many of the same tactics and strategies utilized by brand managers and marketing. For example, creating taglines for a group illustrates its significance (e.g., "You don't have to be lonely at Farmersonly.com"). Another technique is to share testimonials for members who have experienced dramatic positive benefits from their membership. And these testimonials don't have to be life changing, particularly in the case of hobbyist or gaming enthusiasts-type communities. The testimonials may be little more than learning how to defeat one of the most difficult bosses in an online, role-playing game. The point is that that the significance of the accomplishment is important to the community members. This is also the case for another strategy, which is to utilize "celebrity endorsers" and to include what are called "influentials" in the community. If we return to the concepts of nodes from networking theory, then influentials are a particular type of node in the network and it is important to include them in a community. If you think of each individual member of a community as a node, you might find that the average member has three or four ties to others in the community. By way of contrast, influentials will have 12 to 15 ties. Most of those ties will be weak ties, but they can be exploited nonetheless. For example, there may be a particular blogger who has a large following among members of your community. These followers may all be weak ties; however, if as a community manager you can convince that particular blogger to become an active, vocal member of your community, then the other members who already value the blogger's significance will experience the same sense of *gravitas* about the community. Alternatively, assuming that you can't identify influentials for your community, then another method that can be used to enhance a members' experience of significance is to "celebrate the elders" in your community. Chances are that the elders in a community have already established strong name recognition with other members and finding ways to celebrate those elders and to call attention to their participation is has the same effect as celebrity or influential endorsements. The ways that community managers can enhance the sense of significance of experience of their members is only limited by their imagination, but it's important for the long-term success of the community to find ways to promote the value a community has to offer.

Conclusion

In conclusion, we've seen that there are many significant reasons that might lead an enterprise organization to invest in online communities and social networks. They can have a significant impact on the intellectual capital and creative potential of employees. However, we've also seen that the failure to distinguish between the types of network ties upon which the social networks or communities or virtual teams are based can inhibit the types of activities groups can accomplish. If community managers and designers want to set up their social networks and communities for success, it's important that they choose the appropriate networking infrastructure. Finally, once an organization has designed an architecture that will support the viability of its online community, we've examined management techniques: (a) based on lifecycles of membership, (b) based on applying traditional face-to-face management techniques, (c) based types of individual needs of members, and (d) finally based on the lived experiences of members. A knowledge of these fundamentals of online community and social network design should enhance the viability of virtual teams in any enterprise organization.

References

Burt, R. (2000). The network structure of social capital. *Research in Organizational Behaviour*, *22*, 345–423. Retrieved from http://www.bebr.ufl.edu/sites/default/files/The%20Network%20Structure%20of%20Social%20Capital.pdf

Gartner says 80 percent of social business efforts will not achieve intended benefits through 2015. (2013, January 29). *Newsroom* [Press release]. Retrieved from http://www.gartner.com/newsroom/id/2319215

Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, *78*(6), 1360–1380.

Hill, W. F., & Gruner, L. (1973). A study of development in open and closed groups. *Small Group Behavior*, *4*(3), 355–381.

Hinchcliffe, D. (2008, September 8). *Ten leading platforms for creating online communities. ZDNet.* Retrieved from http://www.zdnet.com/article/ten-leading-platforms-for-creating-online-communities/

Howard, T. W. (2010). *Design to thrive: Creating social networks and online communities that last.* Burlington, MA: Morgan Kaufmann.

Just over half of employers using social media tools for internal communication, Towers Watson Survey Finds. (2013, May 23). Retrieved from https://www.towerswatson.com/en-US/Press/2013/05/just-over-half-of-employers-using-social-media-tools-for-internal-communication

Krackhardt, D. (1992). The strength of strong ties: The importance of *philos* in organizations. In N. Nohria & R. Eccles (Eds.), *Networks and organizations: Structure, form, and action* (pp. 216–239). Boston, MA: Harvard Business School Press.

Li, C. (2015, April 7). Why no one uses the corporate social network. *Harvard Business Review.* Retrieved from https://hbr.org/2015/04/why-no-one-uses-the-corporate-social-network

Li, C., & Bernoff, J. (2008). *Groundswell: Winning in a world transformed by social technologies.* Boston, MA: Harvard Business Press.

Kim, A. J. (2000). *Community Building on the Web.* Berkeley, CA: Peachpit Press.

Locke, L. (2007, July 17). The future of Facebook. *Time.* Retrieved from http://content.time.com/time/business/article/0,8599,1644040,00.html

McMillan, R. (2010, March 11). Security industry faces attacks it cannot stop. Network World. Retrieved from http://www.networkworld.com/article/2204266/security/security-industry-faces-attacks-it-cannot-stop.html

Preece, J. (2000). *Online communities: Designing usability, supporting sociability.* Chichester: John Wiley & Sons, Ltd.

Shirky, C. (2008). *Here comes everybody: The power of organizing without organizations.* New York, NY: Penguin Press.

Tuckman, B. W. (1965). Developmental sequence in small groups. *Psychological Bulletin*, *63*, 384–399.

Wenger, E. (1998). *Communities of practice: Learning, meaning, and identity.* Cambridge: Cambridge University Press.

38

Human-Computer Interaction and Education: Designing for Technology-Enhanced Learning Experiences

June Ahn and Tamara Clegg

Introduction

The application of human-computer interaction (HCI) in the domain of education is one of the most exciting areas to explore in our time. The field marries together the possibilities of designing new technologies but in the service of a ubiquitous and universal, human experience: *learning*. We experience learning everywhere (e.g., school, homes, museums, parks etc.), through a variety of experiences (e.g., formal study, hobbies, play etc.), with social peers and mentors (e.g., teachers, students, families, and friends), and in ways that permeate all aspects of our lives (Azevedo, 2013; Bransford, Brown, & Cocking, 2000; Ito et al., 2010; Papert, 1980; Takeuchi & Stevens, 2011). Thus, the design space is complex and the issues surrounding how people learn and in what settings present a rich array of opportunities for designers and researchers.

In this chapter, we provide an overview of how HCI intersects with issues in the field of education. We first explore how the process of design, which is a fundamental component of HCI work, has been conceptualized and used in the field of education. Second, we examine the question of theory and how different theoretical perspectives influence the design of learning technology. For example, cognitive theories of learning focus on how individuals process information and store knowledge in long-term memory. Social and cultural theories of learning focus on how social interaction and culture describe the learning that occurs in a given setting. We highlight how perspectives in HCI mirror these juxtapositions as well, from studies that examine how people process and react to interface changes on the screen, to research that examines the social issues surrounding the use of technology in different contexts. We argue that a key feature of HCI and Education work is to marry design processes with relevant learning theories that can inform richer designs of learning technology. Finally, we build from this framework to suggest ways to conceptualize design for learning.

The Wiley Handbook of Human Computer Interaction, Volume 2, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

Thinking about Design in Education

Design is a fundamental component of work in HCI, and understanding the process of design is a major focus in the field. A major goal in the design process is the creation of a new artifact, product, or tool. However, in the HCI literature, scholars have articulated how *design through research* is also a distinct way of understanding the world (Zimmerman, Forlizzi, & Evenson, 2007). The idea is that the act of creating new innovations is a particular form of research, where the designer must understand something about the world in order to create a tool that solves a unique problem. A focus on design also has led to the creation of *design strategies and methods*. For example, researchers have developed and studied ways to help designers in the process of ideation, iterative development, and testing of new technologies (e.g., Guha et al., 2004; Walsh et al., 2010). Design researchers have developed a variety of techniques to better codesign new technologies with key audiences such as children (Druin, 1999; Guha, Druin, & Fails, 2013; Yip et al., 2013), families (Hutchinson et al., 2003; Walsh, Foss, Yip, & Druin, 2013; Yip et al., 2016), and community stakeholders (Carroll & Rosson, 2007; C. DiSalvo, Nourbakhsh, Holstius, Akin, & Louw, 2008). In addition, scholars have explored how our theories of how things work in the world, and the values that we assume in our theories, greatly impact our designs and the features that end up being created in new technologies (Bardzell & Bardzell, 2011; Friedman & Hendry, 2012; Soloway et al., 1996). These ideas in the HCI field link directly to issues of designing for learning.

In the field of education, and learning sciences, design is conceptualized in slightly different ways. The most common use has been in the area of design-based research or DBR (Collins, Joseph, & Bielaczyc, 2004; Hoadley, 2002). There are many conceptualizations of what DBR is as a methodology (Easterday, Rees Lewis, & Gerber, 2014) and contested debates about how the method is rigorous or useful (Dede, 2004). However, as a general orientation, DBR focuses on iterative cycles of designing a new learning environment or tool, and then using a variety of research methods—both qualitative and quantitative—to better understand how learning occurs with those designed artifacts (Bell, 2004). The results of this cycle should feed into continued, and iterative, refinements of a design, and theories about learning, and then lead to further study. Design-based research has become increasingly popular as a methodology for research in education and the learning sciences (Anderson & Shattuck, 2012), and scholars have written about the need to go deeper into various aspects of the process. For example, Hoadley (2002) advocated early on for richly, descriptive design narratives about the choices that researchers and designers made in the creation of a new technology or tool. He also articulated the differing forms of rigor that characterize DBR, and the potential benefits such as closely linking the design and research process to usable knowledge in complex learning settings (Hoadley, 2004).

In recent works, scholars in both HCI and learning sciences have been increasingly converging. For example, HCI researchers are more strongly advocating for an appreciation for theory, values, and critical frameworks in the design process (Bardzell & Bardzell, 2011; DiSalvo et al., 2008; Friedman & Hendry, 2012; Soloway et al., 1996). Conversely, learning scientists have been increasingly attuned to the intricacies of the design process itself, not merely as an afterthought to get to the research component of most studies (Easterday et al., 2014), as well as applying methods such as participatory design to education research (DiSalvo & DiSalvo, 2014; Yip et al., 2013).

Finally, researchers and scholars are increasingly adopting design methodologies and orientations in other fields such as education leadership and policy implementation. For example, design-based implementation research (DBIR) is an adaptation of DBR that focuses on bringing researchers and education leaders together to undertake iterative cycles of educational change or reform (Penuel, Fishman, Cheng, & Sabelli, 2011). In DBIR, there is a shift in focus from traditional notions where researchers translate and give their insights to practitioners to implement, to developing *joint work* to develop research questions and plans for analysis (Penuel, Allen, Coburn, & Farrell, 2015). Participatory design processes become vital as researchers and practitioners work together to define problems, allot resources, plan research, and conduct evaluations to inform key problems of practice. Likewise, there is increasing adoption of different participatory design methods in domains ranging from developing new curricular materials (Severance, Penuel, Sumner, & Leary, 2016) to working with teachers and education leaders to develop improvements to teaching practices (Cobb, Jackson, Smith, Sorum, & Henrick, 2013). Finally, other research movements in education such as the use of improvement science also foreground the idea that codesigning systemic interventions and related technologies, in partnership with school districts and educators is a pivotal way to engender positive change in education systems (Bryk, 2015).

We observe three synergies between design fields such as HCI and a domain such as education. First, as we noted earlier, design fields often focus on imagining practical and novel solutions for problems. Relatedly, scholars in education have argued that rigorous research in the field should not only be in the form of abstract ideas, but in the relevance and impact of ideas on fostering actual improvement in teaching and learning (Gutiérrez & Penuel, 2014). Second, education researchers often take for granted that the phenomena that they study—be it a new curriculum, technology, or policy—is a designed artifact and that the design itself plays a large role in what consequences we might observe. Attending to issues of usability and how individuals adopt new tools are concepts that are native to HCI and design, and would greatly enhance education research. Third, HCI researchers often focus on the design process of creating new tools. However, HCI researchers often face challenges in deeply understanding the domain of education, teaching, and how people learn. Yet, this understanding becomes important in designing technologies that will actually make an impact. That is, theory from education research plays a vital role in creating usable and effective learning tools.

How Theory Influences Design: Cognitive and Social

Shneiderman and colleagues noted several major trends and critical needs in the field of HCI as of 2016. One of their major observations was a growing shift in design for the individual and technology (micro-HCI) to a turn towards designing for collective experiences (macro-HCI) and thinking about holistic systems when designing new technology (Shneiderman et al., 2016). This commentary reflects a general evolution of HCI that is also seen in education and design research, which builds from individual models of human cognition and interaction to social, cultural, and institutional frameworks. Here we provide a few examples of how strong theoretical understanding in these various domains has a major impact on how one designs new tools for education.

What Shneiderman et al. (2016) describe as micro-HCI maps on well to individual, cognitive theories of learning that have influenced technology design greatly. One illustration involves theories of multimedia learning and cognitive load. Theories of multimedia learning provide robust, evidence-based suggestions for the design of multimedia on a screen (Mayer, 2005; Mayer & Moreno, 2003). For example, researchers have shown that when presenting visual and audio information via multimedia, issues of cognitive load play a major role in how well people can learn from a given interface. Thus, design decisions such as presenting interesting but extraneous information or graphics in a multimedia presentation hinder learning retention for the viewer. Long segments of multimedia presentation (e.g., long video) is less effective than shorter segments that a learner can control by repeating or returning to material, or jumping ahead based on their needs. Furthermore, design decisions such as presenting text or narration, and then presenting visual representations separately on another screen has proven to overload cognition, and thus showing relevant visual representations with explanation together is more effective (Mayer & Moreno, 2003). Human-computer interaction researchers have a long history of adopting psychological and cognitive science perspectives to design for microinteractions—between a human and a computer interface—and have resulted in tremendous new interaction designs.

Another example of how psychological theories inform HCI for education domains comes from theories of scaffolding (Guzdial, 1994; Quintana et al., 2004; Quintana, Zhang, & Krajcik, 2005; Soloway et al., 1996). The idea of scaffolds comes from the observation that learners have prior knowledge and abilities, and they have certain needs for assistance or guidance to help them progress to more complex learning. Some forms of guidance may be too difficult and inaccessible, leading to obstacles for the learner. Other forms of guidance are too easy and thus do not help the learner progress. What learning designers (from a scaffolding perspective) seek is that sweet spot, which Vygotsky termed the zone of proximal development (John-Steiner & Mahn, 1996; Vygotski, 1980), where just enough guidance is given so the learner is challenged but able to progress.

Theories of scaffolding have tremendous implications for interface design. For example, designers will want to embed help and information for users, just in time, and at the right moments of interaction to promote scaffolded learning (Guzdial, 1994). Other researchers have developed tools that help learners document their thought processes, so that they can keep track of their progress, or so peers or teachers might use these data points to provide the right kinds of guidance (Quintana et al., 2004). Knowing how to give guidance, under what conditions, and for different types of learners have direct design implications for learning software (and for software in general).

Subsequently, social and cultural theories about learning and interaction have become prevalent and substantially shape the design of new technologies for society and for learning more specifically. As one example, the theory of distributed cognition has shaped approaches in HCI over the last two decades (e.g., Hollan, Hutchins, & Kirsh, 2000; Nardi, 1996) and has also had a significant impact on learning sciences approaches and technologies as well. Distributed cognition asserts that the active cognitive work people do to carryout tasks and accomplish goals is distributed throughout the social and material space (Hutchins, 1995). For example, pilots flying an airplane distribute tasks between one another (i.e., copilots) and the assortment of materials and tools in the physical space that visually indicate and track the plane's state in relation to important decisions that need to be made (Hutchins, 1995; Hutchins & Klausen, 1996). This theory suggests that instead of viewing learning and cognition as processes solely in an individual's mind, we must acknowledge the ways these processes are shared across people and the environment.

Thus, there has been a move beyond thinking of technology as a means to support individuals, and towards thinking of ways in which technology can distribute the cognitive load between people and the environment. Learning sciences researchers often leverage this approach to design technology to support learning distributed across people. For example, Knowledge Forum is a Web-based system to support communities of learners working together to extend the knowledge base of the entire group (Scardamalia, 2002; Scardamalia & Bereiter, 2006). The system is designed to allow the group to contribute diverse perspectives, information, and insights of community members into a collective whole and to organize such contributions into a collective repository that represents the community's understanding. Other similar approaches have leveraged wiki tools (e.g., Wikipedia) to facilitate such community-based learning (Forte & Bruckman, 2006).

The efforts of recent HCI and learning sciences researchers to support learners' *nomadic* inquiry—or inquiry pursuits distributed across physical spaces and contexts—also has its roots in distributed cognition. This approach seeks to distribute learning throughout the environment as researchers aim to develop new technologies that establish everyday informal contexts as living laboratories. For example, *Zydeco*, a mobile science inquiry app for middle-school learners (Cahill et al., 2010; Kuhn et al., 2012) supports learners' data collection, analysis, and claim development as they move between classroom and informal contexts (e.g., museums). Our own research leverages social media approaches to support learning and inquiry across an even broader range of contexts of learners' everyday lives (i.e., home, school, afterschool, and community settings). The idea is to enable learners to capture and share scientific thoughts, moments, information, and data as they move from place to place—thus enabling them to distribute their learning community (of people) and learning opportunities across settings (Ahn et al., 2014). Others have developed ubiquitous measurement tools that enable learners to collect data in the natural environment (Rogers & Price, 2008). All of these approaches seek to embed inquiry in a range of environments to support learning across contexts.

In addition to cognitive and social perspectives of learning, embodied learning perspectives have recently seen a surge in popularity and increasing influence on the design of new technologies. Embodied learning perspectives posit that learning is inherently connected to one's physical body and the environment (e.g., Barsalou, 1999; Lakoff & Johnson, 1999; Lee, 2015). Whereas previous text-based technologies were less able to leverage embodied experiences for learning, recent advances in sensor-based technologies and wearables enable powerful new types of embodied learning experiences (Lee, 2015). For example, room-sized sensor-based simulations enable learners to become immersed in models of scientific phenomena. In Meteor (Lindgren & Johnson-Glenberg, 2013), for example, learners are asteroids in space, launching their bodies to project their asteroids through space. Wearable technologies also enable new forms of embodied learning experiences (Bower & Sturman, 2015; Lee, 2015). For example, off-the-shelf fitness trackers can be used to help children investigate physical health topics and conduct mathematical analysis relevant to their everyday lives (Carter-Ching & Schaefer, 2015; Lee, Drake, Cain, & Thayne, 2015). Live physiological sensing and visualization tools enable learners to see in real time their physiological body data (e.g., heart rate, breathing rate) on e-textiles and large displays to promote children's content learning about the body and science inquiry practices (Kang et al., 2016; Norooz et al., 2016).

Designing Experiences and Interactions for Learning

In this chapter, we outline how the fields of HCI, education, and learning sciences have brought different facets of design to the project of creating new technologies for learning. For future designers and researchers, there is a rich array of practices that one must develop expertise in to effectively create technologies to teach and learn. On the one hand, there is a great need to develop understanding of different design thinking approaches, and build a toolkit of design strategies that one can employ to create effective user experiences. On the other hand, designers and researchers also need a deep grasp of different learning theories that range from cognitive to social and cultural, to finely specify how one thinks learning occurs and why a newly designed tool will enhance that process.

We end this chapter by arguing that design for learning technologies must also take context and activity structures into account. Education researchers have long observed that learning technologies often do not change teaching and learning practices in formal school settings all that much (Cuban, 1986). Learning scientists have also observed that new technologies need to be designed in a way that aligns with the complex institutions and cultures of school settings (Fishman, Marx, Blumenfeld, Krajcik, & Soloway, 2004). Thus, researchers have documented how technology design alone does not necessarily foster changed behaviors or learning gains. New technologies need to be designed with finely specified and intended *activities* that one posits to lead to effective and engaging learning behaviors (Roschelle, Knudsen, & Hegedus, 2010). New technology can facilitate and enable innovative learning experiences. However, if their use is not well thought out and iterated upon, the effects will be extremely limited (Ahn et al., 2014; Lee et al., 2015; Norooz et al., 2016).

Design-based researchers have thus begun to leverage HCI processes and techniques (e.g., user-centered design, participatory design) to develop learning activities and guidelines for learning activities with these technologies. For example, in our own research, we have explored the design and use of social media technologies to help children engage in scientific inquiry across their life contexts. We are developing a Science Everywhere social media app and large community displays. We had to conduct extensive work in different contexts (i.e., home-, school-, and community-based programs) to understand and develop practices and experiences to promote children's science learning across contexts. We engaged families (i.e., parents and children), teachers, and informal educators as participatory designers of experiences and practices with the technology (Yip et al., 2016). Likewise, we are doing extensive data collection and analysis to understand how the design of the technology and integrated learning experiences promote youth's scientific inquiry across contexts. To promote effective use and scientific inquiry practices with the Science Everywhere technology, we took the approach of using design practices not only to create a social media tool for children's learning but also to understand different stakeholders and the interactions and experiences they have in different learning contexts. The goal is to deeply understand and integrate *practices* in the environment with the new technologies we imagine will enhance those practices. As we move forward with new technologies that transform what is possible with respect to learning, careful, considerate and integrated design and analysis procedures for learning experiences will be critical to realize the future of learning.

References

Ahn, J., Clegg, T., Yip, J., Bonsignore, E., Pauw, D., Gubbels, M.,…Rhodes, E. (2014). Seeing the unseen learner: designing and using social media to recognize children's science dispositions in action. *Learning, Media and Technology*, *41*(2), 1–31.

Anderson, T., & Shattuck, J. (2012). Design-based research a decade of progress in education research? *Educational Researcher*, *41*(1), 16–25.

Azevedo, F. S. (2013). The tailored practice of hobbies and its implication for the design of interest-driven learning environments. *Journal of the Learning Sciences*, *22*(3), 462–510.

Bardzell, S., & Bardzell, J. (2011). Towards a feminist HCI methodology: Social science, feminism, and HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 675–684). New York, NY: ACM. doi:10.1145/1978942.1979041

Barsalou, L. W. (1999). Perceptions of perceptual symbols. *Behavioral and Brain Sciences*, *22*(04), 637–660.

Bell, P. (2004). On the theoretical breadth of design-based research in education. *Educational Psychologist*, *39*(4), 243–253.

Bower, M., & Sturman, D. (2015). What are the educational affordances of wearable technologies? *Computers and Education*, *88*, 343–353.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn*. Washington, DC: National Academy Press. Retrieved from http://www.csun.edu/~SB4310/How%20People%20Learn.pdf

Bryk, A. S. (2015). 2014 AERA distinguished lecture accelerating how we learn to improve. *Educational Researcher*, *44*(9), 467–477.

Cahill, C., Kuhn, A., Schmoll, S., Pompe, A., & Quintana, C. (2010, June). Zydeco: Using mobile and web technologies to support seamless inquiry between museum and school contexts. In *Proceedings of the Ninth International Conference on Interaction Design and Children* (pp. 174–177). New York, NY: ACM.

Carroll, J. M., & Rosson, M. B. (2007). Participatory design in community informatics. *Design Studies*, *28*(3), 243–261. doi:10.1016/j.destud.2007.02.007

Carter Ching, C., & Schaefer, S. (2015). Identities in motion, identities at rest: Engaging bodies and minds in fitness gaming research and design. In V. Lee (Ed.), *Learning technologies and the body: Integration and implementation in formal and informal learning environments* (pp. 201–219). Hoboken, NJ: Taylor & Francis.

Cobb, P., Jackson, K., Smith, T., Sorum, M., & Henrick, E. (2013). Design research with educational systems: Investigating and supporting improvements in the quality of mathematics teaching and learning at scale. *National Society for the Study of Education Yearbook*, *112*(2), 320–349.

Collins, A., Joseph, D., & Bielaczyc, K. (2004). Design research: Theoretical and methodological issues. *The Journal of the Learning Sciences*, *13*(1), 15–42. doi:10.1207/s15327809jls1301_2

Cuban, L. (1986). *Teachers and machines: The classroom use of technology since 1920*. New York, NY: Teachers College Press.

Dede, C. (2004). If design-based research is the answer, what is the question? A commentary on Collins, Joseph, and Bielaczyc; diSessa and Cobb; and Fishman, Marx, Blumenthal, Krajcik, and Soloway in the JLS special issue on design-based research. *The Journal of the Learning Sciences*, *13*(1), 105–114.

DiSalvo, B., & DiSalvo, C. (2014). Designing for democracy in education: Participatory design and the learning sciences. In *Proceedings of the Eleventh International Conference of the Learning Sciences (ICLS 2014)*. Boulder, CO: International Society of the Learning Sciences.

DiSalvo, C., Nourbakhsh, I., Holstius, D., Akin, A., & Louw, M. (2008). The Neighborhood Networks Project: A case study of critical engagement and creative expression through participatory design. In *Proceedings of the Tenth Anniversary Conference on Participatory Design 2008* (pp. 41–50). Indianapolis, IN: Indiana University.

Druin, A. (1999). Cooperative inquiry: developing new technologies for children with children. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: The CHI is the Limit* (pp. 592–599). New York: ACM.

Easterday, M., Rees Lewis, D., & Gerber, E. (2014). Design-based research process: Problems, phases, and applications. In *Proceedings of the International Conference of Learning Sciences* (vol. 14, pp. 317–324). Boulder, CO: International Society of the Learning Sciences.

Fishman, B., Marx, R. W., Blumenfeld, P., Krajcik, J., & Soloway, E. (2004). Creating a framework for research on systemic technology innovations. *Journal of the Learning Sciences*, *13*(1), 43–76. doi:10.1207/s15327809jls1301_3

Forte, A., & Bruckman, A. (2006). From Wikipedia to the classroom: Exploring online publication and learning. In *Proceedings of the Seventh International Conference of the Learning sciences* (pp. 182–188). Bloomington, IN: International Society of the Learning Sciences.

Friedman, B., & Hendry, D. (2012). The envisioning cards: A toolkit for catalyzing humanistic and technical imaginations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1145–1148). New York, NY: ACM. doi:10.1145/2207676.2208562

Guha, M. L., Druin, A., Chipman, G., Fails, J. A., Simms, S., & Farber, A. (2004). Mixing ideas: A new technique for working with young children as design partners. In *Proceedings of the 2004 Conference on Interaction Design and Children: Building a Community* (pp. 35–42). New York, NY: ACM. doi:10.1145/1017833.1017838

Guha, M. L., Druin, A., & Fails, J. A. (2013). Cooperative Inquiry revisited: Reflections of the past and guidelines for the future of intergenerational co-design. *International Journal of Child-Computer Interaction*, *1*(1), 14–23. doi:10.1016/j.ijcci.2012.08.003

Gutiérrez, K. D., & Penuel, W. R. (2014). Relevance to practice as a criterion for rigor. *Educational Researcher*, *43*(1), 19–23.

Guzdial, M. (1994). Software-realized scaffolding to facilitate programming for science learning. *Interactive Learning Environments*, *4*(1), 001–044. doi:10.1080/1049482940040101

Hoadley, C. (2002). Creating context: Design-based research in creating and understanding CSCL. In *Proceedings of the Conference on Computer Support for Collaborative Learning: Foundations for a CSCL Community* (pp. 453–462). New York, NY: ACM.

Hoadley, C. M. (2004). Methodological alignment in design-based research. *Educational Psychologist*, *39*(4), 203–212.

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: Toward a new foundation for human-computer interaction research. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *7*(2), 174–196.

Hutchins, E. (1995). How a cockpit remembers its speeds. *Cognitive Science*, *19*(3), 265–288.

Hutchins, E., & Klausen, T. (1996). Distributed cognition in an airline cockpit. In Y. Engestrom & D. Middleton (Eds.), *Cognition and Communication at Work* (pp. 15–34). Cambridge: Cambridge University Press.

Hutchinson, H., Mackay, W., Westerlund, B., Bederson, B. B., Druin, A., Plaisant, C.,… Eiderbäck, B. (2003). Technology probes: Inspiring design for and with families. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 17–24). New York, NY: ACM. doi:10.1145/642611.642616

Ito, M., Baumer, S., Bittanti, M., boyd, d., Cody, R., Herr-Stephenson, B.,…Tripp, L. (2010). *Hanging out, messing around, and geeking out: Kids living and learning with new media*. Cambridge, MA: MIT Press.

John-Steiner, V., & Mahn, H. (1996). Sociocultural approaches to learning and development: A Vygotskian framework. *Educational Psychologist*, *31*(3–4), 191–206.

Kang, S., Norooz, L., Oguamanam, V., Plane, A., Green, A., Clegg, T., & Froehlich, J. E. (2016). SharedPhys: Live physiological sensing, whole-body interaction and large-screen visualizations to support shared inquiry experiences (pp. 275–287). In *Proceedings of the 15th International Conference on Interaction Design and Children*. New York, NY: ACM.

Kuhn, A., McNally, B., Schmoll, S., Cahill, C., Lo, W. T., Quintana, C., & Delen, I. (2012). How students find, evaluate and utilize peer-collected annotated multimedia data in science inquiry with zydeco. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 3061–3070). New York, NY: ACM.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to Western thought*. New York, NY: Basic Books.

Lee, V. R. (2015). *Learning technologies and the body: Integration and implementation In formal and informal learning environments*. Hoboken, NJ: Taylor & Francis.

Lee, V. R., Drake, J. R., Cain, R., & Thayne, J. (2015). Opportunistic uses of the traditional school day through student examination of Fitbit activity tracker data. In *Proceedings of the 14th International Conference on Interaction Design and Children* (pp. 209–218). New York, NY: ACM.

Lindgren, R., & Johnson-Glenberg, M. (2013). Emboldened by embodiment six precepts for research on embodied learning and mixed reality. *Educational Researcher*, *42*(8), 445–452.

Mayer, R. E. (2005). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 183–200). New York, NY: Cambridge University Press.

Mayer, R. E., & Moreno, R. (2003). Nine ways to reduce cognitive load in multimedia learning. *Educational Psychologist*, *38*(1), 43–52.

Nardi, B. A. (1996). Studying context: A comparison of activity theory, situated action models, and distributed cognition. In B. A. Nardi (Ed.), *Context and consciousness: Activity theory and human-computer interaction* (pp. 69–102). Cambridge, MA: MIT Press.

Norooz, L., Clegg, T. L., Kang, S., Plane, A., Oguamanam, V., & Froehlich, J. E. (2016). “That’s your heart!”: Live physiological sensing and visualization tools for personally relevant and collaborative STEM learning. In C.-K. Looi, U. Cress, J. L. Polman & P. Reimann (Eds.), *Twelfth International Conference of the Learning Sciences* (pp. 779–782). Singapore: International Society of the Learning Sciences.

Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. New York, NY: Basic Books.

Penuel, W. R., Allen, A.-R., Coburn, C. E., & Farrell, C. (2015). Conceptualizing research–practice partnerships as joint work at boundaries. *Journal of Education for Students Placed at Risk (JESPAR)*, *20*(1–2), 182–197.

Penuel, W. R., Fishman, B. J., Cheng, B. H., & Sabelli, N. (2011). Organizing research and development at the intersection of learning, implementation, and design. *Educational Researcher*, *40*(7), 331–337. doi:10.3102/0013189X11421826

Quintana, C., Reiser, B. J., Davis, E. A., Krajcik, J., Fretz, E., Duncan, R. G.,…Soloway, E. (2004). A scaffolding design framework for software to support science inquiry. *The Journal of the Learning Sciences*, *13*(3), 337–386.

Quintana, C., Zhang, M., & Krajcik, J. (2005). A framework for supporting metacognitive aspects of online inquiry through software-based scaffolding. *Educational Psychologist*, *40*(4), 235–244. doi:10.1207/s15326985ep4004_5

Rogers, Y., & Price, S. (2008). The role of mobile devices in facilitating collaborative inquiry in situ. *Research and Practice in Technology Enhanced Learning*, *3*(3), 209–229.

Roschelle, J., Knudsen, J., & Hegedus, S. (2010). From new technological infrastructures to curricular activity systems: advanced designs for teaching and learning. In M. Jacobson & P. Reimann (Eds.), *Designs for Learning Environments of the Future*. Boston, MA: Springer.

Scardamalia, M. (2002). Collective cognitive responsibility for the advancement of knowledge. *Liberal Education in a Knowledge Society*, *97*, 67–98.

Scardamalia, M., & Bereiter, C. (2006). Knowledge building: Theory, pedagogy, and technology. In K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (pp. 97–118). New York, NY: Cambridge University Press.

Severance, S., Penuel, W. R., Sumner, T., & Leary, H. (2016). Organizing for teacher agency in curricular co-design. *Journal of the Learning Sciences*, *25*(4), 531–564.

Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S., Elmqvist, N., & Diakopoulos, N. (2016). Grand challenges for HCI researchers. *Interactions*, *23*(5), 24–25.

Soloway, E., Jackson, S. L., Klein, J., Quintana, C., Reed, J., Spitulnik, J., … Scala, N. (1996). Learning theory in practice: Case studies of learner-centered design. In *Proceedings of the SIGCHI conference on Human factors in computing systems: Common ground* (pp. 189–196). New York, NY: ACM.

Takeuchi, L., & Stevens, R. (2011). *The new coviewing: designing for learning through joint media engagement*. New York, NY: The Joan Ganz Cooney Center at Sesame Workshop.

Vygotski, L. S. (1980). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.

Walsh, G., Druin, A., Guha, M. L., Foss, E., Golub, E., Hatley, L.,…Franckel, S. (2010). Layered Elaboration: A New Technique for Co-design with Children. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1237–1240). New York, NY: ACM. doi:10.1145/1753326.1753512

Walsh, G., Foss, E., Yip, J., & Druin, A. (2013). FACIT PD: A framework for analysis and creation of intergenerational techniques for participatory design. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 2893–2902). New York, NY: ACM. doi:10.1145/2470654.2481400

Yip, J., Clegg, T., Ahn, J., Uchidiuno, J. O., Bonsignore, E., Beck, A.,…Mills, K. (2016). The evolution of engagements and social bonds during child-parent co-design. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 3607–3619). New York, NY: ACM. doi:10.1145/2858036.2858380

Yip, J., Clegg, T., Bonsignore, E., Gelderblom, H., Rhodes, E., & Druin, A. (2013). Brownies or bags-of-stuff? Domain expertise in cooperative inquiry with children. In *Proceedings of the 12th International Conference on Interaction Design and Children* (pp. 201–210). New York, NY: ACM. doi:10.1145/2485760.2485763

Zimmerman, J., Forlizzi, J., & Evenson, S. (2007). Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 493–502). New York, NY: ACM. doi:10.1145/1240624.1240704

39

Digital Citizen Science and the Motivations of Volunteers

Charlene Jennett and Anna L. Cox

Introduction

Around the world, thousands of research projects are engaging millions of non-professional scientists ("citizen scientists") in collecting, categorizing, transcribing, or analyzing scientific data (Bonney et al., 2014). This research practice is known as "citizen science"—where members of the public collaborate with professional scientists to conduct scientific research. Citizen science is growing in popularity due to the innovative use of Web and mobile technologies. For example, volunteers might donate their computers' "downtime" to SETI@home, or classify images of galaxies on the Galaxy Zoo website, or log organisms they have spotted using the Project Noah mobile app.

Digital citizen science projects typically exhibit a skewed pattern of participation, where the majority of participants contribute in small quantities. To build further participation in digital citizen science, it is important to research who participates in digital citizen science, and what motivates them to participate at various levels of engagement. Once we understand what motivates different groups of volunteers to participate in digital citizen science, we can work towards improving the design of digital citizen science projects to better fit the requirements of volunteers, thereby maximizing the opportunities for their engagement.

In this chapter we focus on reviewing HCI research studies that help us to answer the question "What motivates volunteers to participate in digital citizen science?" First we present an overview of different kinds of digital citizen science projects. Then we describe and discuss HCI research studies that have explored volunteers' motivations and also the impact of gamification. We hope that these insights will help to give a better understanding of volunteers, as well as helping to inform the design of citizen science websites and mobile apps.

The Wiley Handbook of Human Computer Interaction, Volume 2, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

Digital Citizen Science

Citizen science is not a new phenomenon. The Christmas Bird Count is thought be the earliest citizen science project; it has been run by the National Audubon Society in the United States every year since 1900, and involves groups of people counting birds in their local area. A crucial difference between then and now, however, is the impact of technology. The Internet has substantially increased the profile of citizen science projects, by increasing project visibility, functionality and accessibility (Bonney et al, 2014). The Internet has also enabled many citizen science projects that can only be accomplished online. Haklay (2013) identifies three kinds of digital citizen science projects (also known as "citizen cyberscience"): volunteer computing, volunteer thinking, and participatory sensing.

In *volunteer computing*, volunteers install software on their personal computers to enable projects to make use of processor time that would otherwise be unused. The Berkeley Open Infrastructure for Network Computing (BOINC) platform allows data to be processed for a range of projects. In SETI@home, one of the BOINC projects, volunteers can donate their spare computer processing capacity to help scientists analyze radio data in the search for extraterrestrials. In LHC@home, volunteers donate spare CPU to run physics simulations to improve the design of the Large Hadron Collider and its detectors. In Rosetta@home, volunteers donate spare CPU to run calculations to simulate protein folding, which has implications for understanding human diseases such as HIV, malaria, cancer, and Alzheimer's. In ClimatePrediction.net, volunteers donate spare CPU to run climate model simulations, helping scientists to understand climate change. For a full list of projects visit the BOINC website (https://boinc.berkeley.edu/projects.php).

In *volunteer thinking*, volunteers are engaged at a more active and cognitive level. Volunteers typically visit a website where they are presented with data and they are trained to analyze the data according to a certain research protocol. The Zooniverse platform hosts many volunteer thinking projects. In Galaxy Zoo, volunteers classify images of galaxies as "ellipticals" or "spirals," and this analysis helps scientists to understand how galaxies are formed. In Bat Detective, volunteers listen to audio recordings and classify bat calls. In Old Weather, volunteers transcribe weather information from pages of historical ships' logs; the weather data is used by scientists for climate modelling and the edited logbooks are of interest to naval and family historians. For a full list of Zooniverse projects visit the Zooniverse website (https://www.zooniverse.org/#/projects).

Volunteer thinking projects can also be found in digital humanities research. For example, in Transcribe Bentham, volunteers transcribe unstudied manuscripts written by the philosopher Jeremy Bentham. In Operation War Diary, volunteers annotate war diaries from the First World War. In Ancient Lives, volunteers transcribe ancient Egyptian papyri.

The third kind of digital citizen science activity is *participatory sensing*. Typically, these projects involve volunteers downloading a mobile phone app, which allows them to collect data by utilizing sensors that are already integrated in their mobile phone. These sensors include different transceivers (mobile network, WiFi, Bluetooth), FM and GPS receivers, camera, accelerometer, online compass, and microphone. For example, in Project Noah and iSpot, volunteers collect data about local animal and plant species. In WideNoise, volunteers collect data about local noise levels. In some participatory sensing projects, volunteers are also asked to submit behavioral information. In Mappiness, volunteers rate how happy they feel in various locations. In Errordiary, volunteers tweet about the errors that they experienced that day.

Volunteers' Motivations in Digital Citizen Science

Several HCI research studies have explored what motivates volunteers to participate in digital citizen science. In particular, researchers wanted to understand why do volunteers initially join a project? And why do some volunteers continue to contribute to a project over several months or years, while other volunteers drop out?

Intrinsic interest in science

Survey studies reveal that volunteers have a variety of motivations for participating in digital citizen science projects; however, a consistent finding is that volunteers typically have an intrinsic interest in the scientific topic.

Raddick et al. (2010) interviewed 20 Galaxy Zoo volunteers and uncovered 12 volunteer motives: contributing to scientific research, learning about galaxies, making discoveries, interacting with the community, teaching others, looking at beautiful images, helping, amazement about the vastness of the universe, interest in the Galaxy Zoo project, interest in astronomy, and a general interest in science. In a subsequent survey of a much larger sample, Raddick et al. (2013) surveyed 10,992 Galaxy Zoo volunteers and found that "I am excited to contribute to original scientific research" was the largest primary motivation, selected by 40% of their respondents. To our knowledge this is the largest sample size for a study exploring digital citizen science motivations to date.

Reed, Raddick, Lardner, and Carney (2013) decided to extend this work to investigate the motivations of Zooniverse volunteers who participated in one or more Zooniverse projects. They conducted a digital survey with 199 Zooniverse volunteers and their results revealed three general categories of motivations for participating in the Zooniverse: social engagement, enjoyment from interacting with the website, and positive feelings from helping Zooniverse projects.

Survey studies with other digital citizen science projects reveal similar findings. Nov, Arazy, and Anderson (2011) surveyed 139 Stardust@home volunteers and identified four main motives: collective, intrinsic, identification, and norm oriented. Crowston and Prestopnik (2013) surveyed Happy Match volunteers and identified three main motives: fun, interest in nature activities, and learning about moths.

Similar results can also be observed in the digital humanities. Causer and Wallace (2012) surveyed 101 Transcribe Bentham volunteers and uncovered several motives, including: an interest in history and / or philosophy, being part of something collaborative, interest in the technology and / or crowdsourcing, interest in Jeremy Bentham, addition of new material, competition, and recognition.

These findings have important implications for designers of future citizen science projects. Given that volunteers have to be intrinsically motivated by the topic itself, it is important to target an appropriate audience, and to consider their needs / requirements. For example, a physics project is likely to be more successfully in recruiting and maintaining volunteers if it makes the physics aspect of the project explicit, and if it is advertised via channels with which physics enthusiasts already engage.

Social features help to sustain engagement

As well as exploring the breadth of volunteers' motivations, it is important to consider the temporal dynamics of volunteers' motivations and how motivations can change over time. Rotman et al. (2012) conducted an online survey with 142 volunteers and scientists of ecological citizen science projects, as well as interviews with 11 volunteers. Their results revealed that volunteers' initial interest in ecological citizen science projects stemmed from elements related to egoism, such as personal curiosity or previous engagement in science projects. Following their initial engagement, volunteers' ongoing participation was affected by factors such as recognition, attribution, feedback, community involvement and advocacy.

Tinati et al. (2014) explored the relationship between task completion and participation in digital discussions in the Zooniverse. They found that out of the 250,071 users in their data, only 40.5% had contributed to both classification tasks and online discussions. The overall positive trend suggests that generally those who contribute more classifications also contribute more forum posts. A set of active users were responsible for over 70% of the digital discussion, assuming the role of the "core community." They also identified several roles of volunteers in online discussion forums: general help asker, answerer, informers, moderator, discoverer, hypothesizer, investigator / validator, cheerleader, and celebrator.

Similarly, we conducted a study to investigate how low contributors differed from more committed volunteers in Old Weather (Eveleigh, Jennett, Blandford, Brohan, & Cox, 2014). We distributed a survey to 200 Old Weather volunteers, followed by interviews with 17 respondents selected according to a range of contribution levels. Our findings revealed that high contributors posted over 500 forum posts on average, whereas low contributors chose not to engage in digital discussions (<1 forum post on average). High contributors were deeply engaged by social or competitive features, whereas low contributors described a solitary experience of "dabbling" in projects for short periods.

Given the extensive contributions to community features made by some participants, and the importance that these volunteers place on being able to contribute to the community, it is important that projects provide opportunities for volunteers to communicate with each other (and directly with the scientists). The Zooniverse design team recently reflected on the development of the Zooniverse platform and they confirm that social features and community support are vital to project success (Tinati et al., 2015). Digital citizen science volunteers want and need the ability to discuss aspects of projects and tasks. Timely support from scientists is also crucial. The Zooniverse has even experienced a number of exciting citizen-led discoveries, as a consequence of volunteers going beyond the requirements of the task and asking questions about unusual objects within an observation. For example, Galaxy Zoo's discovery of "green peas" was a result of volunteers observing green galaxies in the images, which scientists had originally thought were just green glitches in the imaging apparatus.

However, although community-oriented volunteers tend to participate more, it is also important to remember that not all volunteers want to be sociable online.

The majority of participants exhibit a small-scale contribution pattern but together can account for a huge percentage of the total contributions made to a project. In addition, there is no evidence that their contributions are less valuable than those from more regular volunteers. We therefore argue that it is just as important to design for dabblers as it is to design features for regular contributors, as there is great potential value in designing interfaces to tempt lone workers to complete "just another page." Our design recommendations include breaking the work into components that can be tackled without a major commitment of time and effort, and providing feedback on the quality and value of these small contributions (Eveleigh et al., 2014).

A further consideration is that are also instances where social factors inhibit participation. We conducted an interview study with eight Errordiary contributors (Jennett et al., 2014). We observed that participants found it "fun" to share their errors with others. Social factors such as "helping" and "sharing" encouraged them to join and contribute to the project. At the same time, however, we found Errordiary participants were sometimes put off posting because they thought others might view their contribution as "mundane" or "not funny." Our findings highlight that one of the main ways that Errordiary, a citizen-psych science project, differs from citizen science projects in other domains (e.g. the physical and biological sciences) is the personal nature of the data. This example illustrates the potential for many nuances in volunteer behaviors as the spectrum of digital citizen science grows. Future research is needed to understand which findings are generalizable, and which findings are domain-specific or project-specific.

Reasons for dropping out

Surveys and interviews are usually filled in by volunteers that are actively involved, which means that the point of view of volunteers that are no longer involved is sometimes missed. In Eveleigh et al. (2014), we were able to recruit several Old Weather volunteers for our interview study that were no longer active on Old Weather but were still subscribed to the mailing list. This allowed us to gain insights into some of the reasons why volunteers drop out. We found that a few volunteers were unsure about the usefulness of their contributions. For example, one participant says "I lost motivation to continue contributing information because I was not sure how useful my input was…" Some volunteers were also concerned that they might be providing inaccurate data. For example, another participant says "I really like the concept but I had trouble deciphering the handwriting. So I was afraid I was getting things wrong…" These findings highlight the importance of providing regular project updates and reassuring volunteers in their ability to contribute to research. Boredom and lack of time were other reasons that volunteers gave for dropping out.

We suggest that there are many ways that projects can inform volunteers about their personal progress and project's progress (Jennett & Cox, 2014). Designers of digital citizen science projects can utilize progress bars (e.g. you have completed 2 out of 4 steps), counters (e.g. you have contributed 6 photos), and project blogs. Gamification mechanisms, such as badges and leaderboards, might also be worth considering—this is discussed more in the next section.

GWAPs and Gamification

The popularity of casual games, and the explosion of Internet-enabled devices—such as laptops, tablets, and smartphones—has helped to ensure that gaming is part of many people's everyday lives. With this in mind, it is not surprising that scientists began to wonder "what would happen if we made citizen science tasks more fun, like a game?" Would this help to motivate volunteers to take part? "Games with a purpose" (GWAPs) are games that are created to be fun, while also encouraging people to solve a problem that computers cannot yet solve (von Ahn, 2006).

Foldit is a GWAP developed by scientists at the University of Washington. Foldit challenges players to predict and design the structures of proteins in the hopes of better understanding how they work. Players can chat live with other players, join teams, and track their progress on leaderboards. In 2011, Foldit appeared in scientific news headlines around the world because a team of players ("Folders") were able to decipher the structure of a protein called retroviral protease in less than 2 weeks. This enzyme is key to the way HIV multiplies and it was a structure that scientists had been trying to solve for the past decade. This discovery established Foldit as a legitimate resource in the study of protein folding. It also showed that games can be used successfully to tap into the wisdom of the crowd (Good & Su, 2011).

Many other citizen science games have been launched online. In Eyewire, players map the connections between neurons. In Cell Slider, players review images to spot cancer cells. In Malaria Hunters, players review images to diagnose malaria. In Phylo, players move color blocks on the screen to come up with new gene combinations. In Quantum Moves, players help to build a quantum supercomputer by simulating laser beams to move atoms onto their correct pathways. In Play to Cure: Genes in Space, players try to collect a fictional substance called "Element Alpha," which represents genetic cancer data. More citizen science games can be found on the Citizen Science Center website (http://www.citizensciencecenter.com/citizen-science games-ultimate-list/)

"Gamification" is a term that refers to the use of game elements in nongame contexts (Deterding, Sicart, Nacke, O'Hara, & Dixon, 2011). It is a broad term that encompasses GWAPs, which typically feel like a game when you play them, but also other kinds of digital citizen science tasks that utilize leaderboards and badges. For example, Old Weather could be viewed as a gamified volunteer thinking project. Old Weather uses a ranking system where the volunteer that transcribes the most pages for a particular ship is awarded the title of "captain" of that ship. This ranking mechanism is designed to encourage volunteers to stay loyal to a particular ship and to increase their familiarity with the handwriting in that logbook, rather than attempting to transcribe random pages of different ships.

Volunteers' Motivations in Gamified Citizen Science

The gamification of citizen science tasks has led to several interesting HCI research questions. Why do volunteers play citizen science games? Are the motivations of citizen science "players" different to volunteers that contribute to non-gamified citizen science tasks? Does gamification appeal to everyone, or only certain groups of people?

Science is still the main interest

Some scientists worry about data quality in citizen science games, as it is possible that players may attempt to "game the system" in the pursuit of points (Bowser, Hansen, & Preece, 2013). Importantly, the results of two interview studies suggest that players do care about the science, and players show similar motivations to volunteers of non-gamified citizen science projects.

In our interview study we interviewed four Foldit players and four Eyewire players to find out more about why they took part in their respective projects (Iacovides, Jennett, Cornish-Trestrail, & Cox, 2013). Like previous citizen science studies (e.g. Rotman et al. 2012), most of our participants had a prior interest in science and had found out about the project through science-related magazine and websites. Essentially, they were not attracted to the project because they were interested in games but because they were interested in science. Game elements did appear useful, however, in helping to sustain participants' engagement in the project. For example, one participant said "the points don't motivate me but they do drive me further." Being part of a team was another factor that encouraged Foldit players to continue participating. Another participant describes "if there were no group I wouldn't be involved."

Curtis (2015) interviewed nine Foldit players and her findings reveal similar motivations. In her thesis she describes how participants felt motivated to participate because they were part of a community and had a shared goal. They also described how Foldit gave them the opportunity to take part in scientific research and make a contribution to science. Again it appears that participants were primarily attracted to the science aspects of the projects and not just because they wanted to play a game. Further research is needed, however, as both sample sizes are small and these results might not generalize to all kinds of citizen science games.

Gamification has mixed appeal

Another important HCI research finding is that gamification does not appeal to everyone. Bowser, Hansen, & Preece (2013) carried out an evaluation of Floracaching, a geocaching game that encourages players to gather plant phenology data. They recruited 58 participants (22 plant experts and 36 technology enthusiasts) and found that some participant did enjoy Floracaching because it was game-like. On the other hand, they also found that hardcore citizen scientists may eschew game-like aspects for a more serious interface. For example, one plant enthusiast who tested Floracaching said that he found the game-like elements distracting and anther advocated for more tools for plant experts, such as a taxonomic key. These results suggest a design challenge in making apps appeal to both citizen scientists and casual gamers.

Similarly, in our interview study with 18 Old Weather volunteers, we found that the ranking system to be "captain" of a ship had mixed appeal (Eveleigh, Jennett, Lynn, & Cox, 2013). Some volunteers felt that the ranking system helped to validate their efforts and they felt motivated to achieve status of captain. For example, one participant said "I was captain of a vessel and it felt rather good. Even though it doesn't mean anything per se, to know that I had achieved more log pages than anyone else." On the other hand, other volunteers described how they found the ranking system to be demotivating, as they felt like they couldn't catch up: "I was never even close to it.

It seemed like you had to transcribe 10 times as much as I was transcribing…" Together these findings suggest that it is important to offer a balanced range of game-like features, so as to encourage the most active contributors, while at the same time still supporting and encouraging new volunteers trying out the project.

Another key finding from our interview study is that transcription projects such as Old Weather are motivating because of their narrative appeal. As one participant describes, "The 'real' story that those logs imply is as hypnotically fascinating as any form of fiction or non-fiction…The ship and crew became friends and even the handwriting became clues as to whom had the watch for that day…" This opens up some interesting possibilities for future research, as it is important to remember that gamification is not limited to just points, badges, and leaderboards. It could be possible to apply game-like features, which promote immersion in stories, such as the emerging story of the project itself as it progresses.

Potential for attracting millennials

Bowser, Hansen, He et al. (2013) suggest that gamification can inspire new citizen science volunteers. They conducted an evaluation with "Biotracker," a gamified mobile app that gather plant phenology data. In their study they recruited participants born after 1980, also known as "millennials" or "digital natives." Millennials use technology more frequently than their elders, play video games more frequently, and generally have more positive attitudes towards technology. During the evaluation, 71 millennials tested out Biotracker and filled in an online survey about their experience. The results revealed that while most millennials may not embrace a gamified citizen science app, a significant portion (14%) are likely to engage with the app because it is gamified. These participants were attracted to elements of gamification (earning badges, competing with peers), social motivations (community membership, socialization), and personal benefits such as fun and education. They did not express any of the typical motivations that are held by traditional citizen science volunteers, such as the desire to contribute to science and the desire to contribute to the public good. This suggests that that gamification is key to attracting millennials and it could help citizen science campaigns to reach new audiences.

On the other hand, Bowser, Hansen, He et al. (2013) also found that millennials may not be as patient with a gamified citizen science as citizen scientists who are already motivated to volunteer their data. Millennials expect that technology should make their life easier, so it becomes especially important to ensure that gamification is pervasive and well designed, and that any usability issues are resolved.

Conclusions

Human-computer interaction research plays a vital role in digital citizen science, as it enables researchers and designers to gain a better understanding of their volunteers and why they participate as often (or as little) as they do. In this book chapter we have described several HCI studies that have explored issues related to volunteers' motivations in various digital citizen science projects. We have also presented several design recommendations based on these findings.

Overall it is evident that volunteers are motivated primarily because they want to contribute to science. Volunteers who contribute long term are often motivated by social features and community aspects of projects. However, not everyone wants to be active in an online community or has time to contribute long term, so it is just as important to design for dabblers as it is to design features for regular contributors.

Regarding gamification, studies show that players of citizen science games are also primarily motivated by the science. We suggest that there is no reason to think that these players are less committed than volunteers of non-gamified projects. However there is potential for gamified projects to attract new audiences, such as millennials, and so this landscape could end up changing. Different designs might need to be put in place to keep both new and old audiences happy.

Considering the scope of research that has been conducted so far, it is evident that existing studies have mostly focused on the experiences of volunteers in volunteer thinking projects, and less is known about the experiences of volunteers in volunteer computing and participatory sensing. There has also been more focus on well-known and successful digital citizen science projects and games (e.g. Galaxy Zoo, Foldit), and less is known about newer projects and those that are less successful. As more and more HCI research is conducted in this area, we hope that more can be uncovered about volunteers' experiences across a range of different kinds of digital citizen science projects.

Websites of Cited Digital Citizen Science Projects

Ancient Lives. http://www.ancientlives.org/
Bat Detective. http://www.batdetective.org/
BOINC. https://boinc.berkeley.edu/projects.php
Cell Slider. www.cellslider.net
Citizen Science Center. http://www.citizensciencecenter.com/citizen-science-games-ultimate-list/
ClimatePrediction.net. http://www.climateprediction.net/
Errordiary. http://www.errordiary.org/
Eyewire. www.eyewire.org
Foldit. https://fold.it/portal/
Galaxy Zoo. http://www.galaxyzoo.org/
Happy Match. http://www.citizensort.org/web.php/happymatch
iSpot. http://www.opalexplorenature.org/ispot
LHC@home. http://lhcathome.web.cern.ch/
Malaria Hunters. www.malariaspot.org
Mappiness. http://www.mappiness.org.uk/
Old Weather. http://www.oldweather.org/
Operation War Diary. http://www.operationwardiary.org/
Phylo. http://phylo.cs.mcgill.ca/
Play to Cure: Genes in Space. http://www.cancerresearchuk.org/support-us/play-to-cure-genes-in-space
Project Noah. http://www.projectnoah.org/
Quantum Moves. http://www.scienceathome.org

Rosetta@home. http://boinc.bakerlab.org/rosetta/
SETI@home. http://setiathome.ssl.berkeley.edu/
Stardust@home. http://stardustathome.ssl.berkeley.edu/
Transcribe Bentham. http://blogs.ucl.ac.uk/transcribe-bentham/
WideNoise. http://cs.everyaware.eu/event/widenoise/
Zooniverse. https://www.zooniverse.org/#/projects

Acknowledgements

Jennett and Cox were funded by the Citizen Cyberlab project (EU FP7 #317705) and the Open 3D project (EPSRC grant EP/M013685/1).

References

Bonney, R., Shirk, J. L., Phillips, T. B., Wiggins, A., Ballard, H. L., Miller-Rushing, A. J., & Parrish, J. K. (2014). Next steps for citizen science. *Science*, *343*(6178), 1436–1437. doi:10.1126/science.1251554

Bowser, A., Hansen, D., He, Y., Boston, C., Reid, M., Gunnell, L., & Preece, J. (2013). Using gamification to inspire new citizen science volunteers. In *Gamification'13: Proceedings of the First International Conference on Gameful Design, Research and Applications* (pp. 18–25). New York, NY: ACM.

Bowser, A., Hansen, D., & Preece, J. (2013, April/May). *Gamifying citizen science: Lessons and future directions.* Paper presented at the Designing Gamification: Creating Gameful and Playful Experiences workshop at CHI'13 conference on Human Factors in Computing Systems, Paris, France.

Causer, T., & Wallace, V. (2012). Building a volunteer community: Results and findings from Transcribe Bentham. *Digital Humanities Quarterly*, *6*(2). Retrieved from http://www.digitalhumanities.org/dhq/vol/6/2/index.html

Crowston, K., & Prestopnik, N. R. (2013). Motivation and data quality in a citizen science game: A design science evaluation. In *Proceedings of the 2013 46th Hawaii International Conference on System Sciences (HICSS'13)* (pp. 450–459). doi:10.1109/HICSS.2013.413

Curtis, V. (2015). Online citizen science projects: an exploration of motivation, contribution and participation (Unpublished PhD thesis). The Open University, Milton Keynes.

Deterding, S., Sicart, M., Nacke, L., O'Hara, K., & Dixon, D. (2011). Gamification: Using game design elements in non-gaming contexts. In *Proceedings of CHI '11: Extended Abstracts on Human Factors in Computing Systems* (pp. 2425–2428). New York, NY: ACM.

Eveleigh, A., Jennett, C., Blandford, A., Brohan, P., & Cox, A. L. (2014). Designing for dabblers and deterring drop-outs in citizen science. In *Gamification'13: Proceedings of the First International Conference on Gameful Design, Research, and Applications* (pp. 2986–2994). New York, NY: ACM.

Eveleigh, A., Jennett, C., Lynn, S., & Cox, A. L. (2013). "I want to be a captain! I want to be a captain!" Gamification in the Old Weather citizen science project. In *Proceedings of Gamification 2013* (pp. 79–82). New York, NY: ACM.

Good, B. M., & Su, A. I. (2011). Games with a scientific purpose. *Genome Biology*, *12*, 135.

Haklay, M. (2013). Citizen science and volunteered geographic information—overview and typology of participation. In D. Z. Sui, S. Elwood, & M. F. Goodchild (Eds.), *Crowdsourcing* *geographic knowledge: Volunteered geographic information (VGI) in theory and practice* (pp. 105–122). Berlin: Springer Science + Business Media. doi:10.1007/978-94-007-4587-2_7

Iacovides, I., Jennett, C. Cornish-Trestrail, C., & Cox, A. L. (2013). Do games attract or sustain engagement in citizen science? A study of volunteers motivations. In *Proceedings of CHI'13: Extended Abstracts on Human Factors in Computing Systems* (pp. 1101–1106). New York, NY: ACM.

Jennett, C., & Cox, A. L. (2014, November). *Eight guidelines for designing virtual citizen science projects.* Paper presented at "Citizen + X: Volunteer-Based Crowdsourcing in Science, Public Health and Government" workshop, at HCOMP'14 conference on Human Computation and Crowdsourcing, Pittsburgh, PA. Retrieved from https://www.aaai.org/ocs/index.php/HCOMP/HCOMP14/paper/view/9261

Jennett, C., Furniss, D., Iacovides, I., Wiseman, S., Gould, S. J. J., & Cox, A. L. (2014). Exploring citizen psych-science and the motivations of errordiary volunteers. *Human Computation*, *1*(2), 200–218.

Nov, O., Arazy, O., & Anderson, D. (2011). Dusting for science: motivation and participation of digital citizen science volunteers. In *Proceedings of iConference 2011* (pp. 68–74). doi:10.1145/1940761.1940771

Raddick, M.J., Bracey, G., Gay, P. L., Lintott, C. J., Cardamone, C., Murrary, P.,…Vandenberg, J. (2013). Galaxy Zoo: Motivations of citizen scientists. *Astronomy Education Review*, *12*(1), 010106. doi:10.3847/AER2011021

Raddick, M. J., Bracey, G., Gay, P. L., Lintott, C. J., Murray, P., Schawinski, K.,…Vandenberg, J. (2010). Galaxy Zoo: Exploring the motivations of citizen science volunteers. *Astronomy Education Review*, *9*(1), 010103-1. doi:10.3847/AER2009036

Reed, J., Raddick, M. J., Lardner, A., & Carney, K. (2013). An exploratory factor analysis of motivations for participating in Zooniverse, a collection of virtual citizen science projects. In *Proceedings of the 46th Hawaii International Conference on System Sciences (HICSS 2013)* (pp. 610–619). doi:10.1109/HICSS.2013.85

Rotman, D., Preece, J., Hammock, J., Procita, K., Hansen, D., Parr, C.,…Jacobs, D. (2012). Dynamic changes in motivation in collaborative citizen-science projects. In *Proceedings of CSCW 2012* (pp. 217–226). New York, NY: ACM.

Tinati, R., Kleek, M. V., Simperl, E., Luczak-Rösch, M., Simpson, R., & Shadbolt, N. (2015). Designing for citizen data analysis: A cross-sectional case study of a multi-domain citizen science platform. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI'15)* (pp. 4069–4078). doi:10.1145/2702123.2702420

Tinati, R., Simperi, E., Luczak- Rösch, M., Kleek, M. V., & Shadbolt, N. (2014, June 10–12). Collective intelligence in citizen science—a study of performers and talkers. Paper presented at Collective Intelligence 2014, Cambridge, MA. Retrieved 5 October, 2017, from the arXiv database https://arxiv.org/abs/1406.7551

von Ahn, L. (2006). Games with a purpose. *Computer*, *39*(6), 92–94. doi:10.1109/MC.2006.196

Part XI
Applications

40

Motor Vehicle Driver Interfaces

Paul A. Green

Introduction

This chapter considers the driver interface, also known as the human-machine interface (HMI), which concerns the controls and displays with which the driver (and sometimes passengers) can interact, as well as the supporting operating logic.

This chapter is written for professionals familiar with human-computer interaction (HCI) in general, but not familiar with its application to motor vehicles. An underlying theme is that the safety-critical and highly regulated nature of driving leads to significant departures from standard HCI practice. Even though the appearance of these interfaces may resemble that of interfaces commonly found on smartphones and tablets, some of the methods, measures, and statistics to evaluate them are unique to motor-vehicle applications. For non-HCI professionals, reading chapters earlier in this text should provide the desired background. For driver interface developers, this chapter should gather together information dispersed throughout the literature.

Human-computer interaction is of interest to motor-vehicle developers because of widespread and growing use of computer interfaces in motor vehicles. Among them are interfaces for operating basic vehicle driving functions, driving assistance systems (e.g., adaptive cruise control, lane-keeping assistance), navigation systems, entertainment systems, and smartphones. Interfaces to support these systems are being developed to (a) enhance driving safety, (b) make transportation more efficient (saving time and fuel), (c) make driving more enjoyable, and (d) make drivers more productive.

Although findings from research are important, this chapter emphasizes the resulting design documents and evaluation methods for driver interfaces to promote safety and ease of use. As this is a reference handbook, engineering practice receives more attention than scientific theory. Furthermore, given its technology focus, the design of traditional (noncomputer) driver interfaces (such as switches for headlights and windshield wipers) is not covered. For information on traditional interfaces, the best source is Bhise's (2011) book, *Ergonomics in the Automotive Design Process*. Readers may also find that Akamatsu, Green, and Bengler (2013) provides a useful perspective.

The Wiley Handbook of Human Computer Interaction, Volume 2, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

In their classic paper on usability, Gould and Lewis (1985) identified three key principles to be followed when designing products for ease of use:

- Early focus on users and tasks.
- Empirical measurement.
- Iterative design.

These principles not only apply to office applications and web development but automotive applications as well. In the automotive context, developers need to understand (a) who drives the vehicle (users), (b) the driving task (the most important task), (c) what in-vehicle tasks users perform, (d) task context, and (e) the consequence of task failures. These topics are the focus of the first part of this chapter.

Second, it is important to be able to measure driver and system performance (empirical measurement). That topic constitutes the second part of this chapter.

Surprisingly, there have been few reports of how iterative design is used in developing driver interfaces, although the approach is used. Complete attention to all three principles, however, is not common (Lee, Forlizzi, & Hudson, 2008; Steinfeld & Tan, 2000). Because a great deal of automotive design relies upon following design standards, that topic is the final focus of this chapter.

What is the Driving Context in which Users Perform Tasks?

According to the World Health Organization (2013), over 1.2 million people die in road-traffic crashes each year, or almost 3,300 per day, and somewhere between 20 and 50 million suffer injuries. WHO ranked traffic crashes as the ninth leading cause of death and the leading cause of death of adults ages 15–29. If the current trends continue, by the year 2030, traffic crashes will become the fifth largest cause of death after heart attacks, stroke, pneumonia, and lung diseases of various types.

Additional insights come from crash data for the United States, for which reliable, detailed crash statistics are available. In fact, the United States is probably the only country in the world for which its crash databases are available to anyone for free, which unfortunately can lead those examining crash data to a U.S.-centric perspective.

Analyses of U.S. crashes typically rely on three databases: (a) the Fatality Analysis Reporting System (FARS), (b) the National Automotive Sampling System (NASS) General Estimates System (GES), and (c) the Crashworthiness Data System (CDS). The FARS (https://www.nhtsa.gov/research-data/fatality-analysis-reporting-system-fars) is a database containing all fatal crashes in the United States. The GES (NHTSA, NASS General Estimates System) is a nationally representative sample of police-reported crashes of all severities (including those that result in death, injury, or property damage). The CDS (NHTSA, Crashworthiness Data System Overview) is an annual probability sample of approximately 5,000 police-reported crashes involving at least one passenger vehicle that was towed from the scene (out of a population of almost 3.4 million tow-away crashes). Minor crashes (involving property damage only) are not in CDS. The CDS crashes are investigated in detail by specially trained teams of professionals who provide much more information than is given in police reports.

According to the U.S. Department of Transportation's 2014 annual traffic-safety assessment (U.S. Department of Transportation, 2015), 32,675 people were killed in traffic crashes in the United States in that year. Of them, about 11,926 were in passenger cars, 9,096 were in light trucks, 657 were in heavy trucks, 44 were in buses, 4,586 were motorcyclists, and 553 were associated with other vehicles. The remaining 5,813 deaths primarily involved pedestrians and bicyclists.

How often and what kinds of crashes are associated with driver interfaces?

Crashes can occur for a wide variety of reasons (U.S. Department of Transportation, 2008) and crashes are often attributable to multiple causes. In the United States, the most commonly cited causes are intoxication and speeding (Kolash, 2014). Driver distraction continues to be an issue that receives considerable attention, with particular concern for the driver interface. There are a number of useful summaries of distracted driving, with a recent U.S. Department of Transportation report being a good starting point (U.S. Department of Transportation, 2016a). Key points that report makes follow.

In 2014, there were 3179 people killed and an estimated additional 431,000 injured in motor vehicle crashes involving distracted drivers. Those totals represent 10% of all fatal crashes, 18% of injury crashes, and 16% of all police-reported motor vehicle traffic crashes in 2014. Drivers ages 15–19 are most likely to be involved in distraction crashes.

When considering the data, realize that those totals represent all sources of distraction, not just smartphones (discussed in detail later), the common association. Totals in previous years may be less because what is considered a distraction may have changed. Note that, in most cases, distraction, as a causal factor, was probably not observed by the police officer writing the crash report, so there can be concerns about the reliability of such data.

There are two phenomena associated with distraction. The first is that the more often drivers look away from the road, the more likely they are to not see a crash-provocative situation and the greater the probability of a crash. For tasks such as entering a street address, this is the problem. Second, some distracting tasks lead to long off-road glance durations because those tasks attract attention. The longer the driver looks away from the road, the longer the individual glance time becomes, and therefore, the less likely the driver is to have current knowledge of the driving situation. The risk function for glance duration is likely to have an exponent greater than 1. What can therefore occur is that drivers become so engaged in the secondary task that they fail to realize how much time has passed since they last looked at the road, and look inside the vehicle for too long. Overall, the product of these two factors is total off-road glance time, a measure of exposure. For most tasks, the number of off-road glances is the primary contributor to crash risk (Liang, Lee, & Yekhshatyan, 2012).

What kind of trips do people make and why?

Every 5 to 10 years, the U.S. Department of Transportation conducts the National Household Travel Survey to obtain travel data for the United States (Santos, McGuckin, Nakamoto, Gray, & Liss, 2011), and many other countries conduct similar studies as well. (See UK Department for Transport, n.d.) In U.S. data from 2009, people in the United States were shown to have traveled an average of 14,500 miles

Table 40.1 Summary of trip purposes.

Purpose	*% person trips*
Family and personal business	44.6
Work	14.8
Social & recreational	27.1
School & church	9.8
Work-related	2.9
Other	0.8

Source: Hu and Reuscher (2004).

per year, making four trips per day. They drove an average of 40 miles per day, with most of the miles (about 35) covered in a personal vehicle. Keep in mind that these are averages, and that public transit (including school buses) prevalent in urban areas accounts for only 2% of all trips. The travel situation is likely to be different for more urbanized countries (Japan, most of Europe), where public transit is more prevalent.

According to the 2001 data (Table 40.1), the most common reason for travel is family and personal business, which includes shopping, running errands, and dropping off and picking up others, accounting for almost half of the trips.

These and other data (on trip distances, travel speeds, time of day, etc.) in the National Household Travel Survey provide information on both the tasks and information needs that driver information systems should support and the conditions (road types, speed, weather, etc.) under which safety and usability should be assessed. For additional information on travel, see Lo, Green, and Franzblau (2011).

Finally, consider that the travel survey data are for personal vehicles. However, a substantial fraction of all miles traveled is by trucks (Davis, Diegel, & Boundy, 2015), and the interfaces described in this chapter will be implemented in them as well.

In contrast to the emerging understanding of the primary driving task, less is known about the real use of in-vehicle devices while driving, in particular the frequency and duration of various tasks, though naturalistic driving studies are beginning to provide insights (Sayer, Devonshire, & Flannagan, 2005).

Who are the users?

Unlike computer users, operators of motor vehicles must be licensed. In the United States, the process of becoming a licensed driver begins with obtaining a copy of the state driving manual and learning the state's traffic laws. Candidates must also pass vision tests (see Low Vision Resources Center, 2003) and take a test of rules of the road to obtain a learner's permit, often on or after their 16th birthdays. Consistent with the increasingly common practice of graduated driver licensing, learners can drive at restricted times with adult supervision. They must generally complete a driver's education class and, after a few years, they pass an on-the-road test and obtain a license to drive. (For details, see Highway Loss Data Institute, n.d. b.) Graduated licensing provides new drivers with more experience under less risky conditions, thereby reducing crash risk. Specifically, Mayhew, Simpson, and Pak (2003) showed that crash rates per 10,000 novice drivers drop dramatically with time, being about 120, 100, and 70 after 1, 3, and 6 months, respectively, of being licensed. Similarly, recognizing the increased risk of elderly drivers, some states have special renewal procedures for older drivers (Highway Loss Data Institute, n.d. a).

In the United States, obtaining a commercial driver's license needed to drive buses, large trucks, and other vehicles is a more complex process. Most candidates either obtain (a) on-the-job training, (b) training integrated into their lifestyle (using machinery on a farm), or (c) training at truck driving schools (Sloss & Green, 2000). That population tends to be older than the working population as a whole, and is predominantly male (Short, 2014).

Driver licensing practice varies from country to country. Europe has had a common license in place since 2013. Japan has had a notably difficult-to-pass licensing process, with a relatively high failure rate for the basic licensing exam even though there is substantial enrollment in special schools to train drivers. In some countries, obtaining a license can be the opposite, requiring minimal skill, training, or knowledge, and corruption of the licensing authority can be an issue (Bertrand, Djankov, Hanna, & Mullainathan, 2008).

Although there are significant differences between countries in crash rates and driving culture, this section will focus on U.S. drivers because the U.S. driving data is the most comprehensive data available. For the United States, the most current overview data on licensed drivers appears in *Highway Statistics*, an annual publication of the Federal Highway Administration. In the United States, within any age group, the percentages of men and women who are licensed are almost equal except for the elderly (U.S. Department of Transportation, 2016b). Elderly women are sometimes more likely to drive because they are in better health than men of the same age. The percentage of the U.S. population that is licensed is 74% at age 21 and increases with age, reaching a maximum at age 60–64 (92%), and then begins to decline. Even for the 85 and older group, 58% are licensed drivers. Thus, in designing in-vehicle systems for motor vehicles, few adults can be excluded, which differs from the design of office computer systems, where the emphasis is on the working population (generally less than 65 years old). Further, because of a wide range of age, skill, and experience, significant differences in individual performance can be expected. For example, in UMTRI driver interface studies, older drivers typically required one-and-a-half to two times longer to complete tasks than younger drivers (Green, 2001c). This fact, along with the requirement *to design and test for the reasonable worst-case drivers, makes including drivers over age 65 imperative.* Of course, this is all for the United States, and in places where the vehicle market is growing rapidly such as China, there is a greater predominance of younger drivers. However, even there, some older adults drive, so excluding them from the user population is inappropriate.

What kinds of vehicles do people drive (the platform question)?

For computers, people are concerned about the (a) brand, (b) amount of memory, (c) processor speed, (d) capacity of the hard drive, (e) type and version of operating system (Windows, Mac, or Linux), (f) the number of monitors and resolution supported by the video card, (g) the network connection speed, and so forth. The hardware and software of individual computers is in a state of flux, being constantly updated over a lifespan of often 3 to 5 years ("Average life of a desktop PC 2012–2017," 2016). In the past, the desktop computer physical interface was fairly consistent—a QWERTY keyboard, mouse, and a large monitor. For many, smartphones and tablets are replacing their desktop computer (Anderson, 2015). The on–screen "desktop" is a more flexible space than the motor vehicle instrument panel, though there has been discussion of personalizing instrument clusters, complete with personal photos (Ziomek, Tedesco, & Coughlin, 2013).

In contrast to personal computers, a motor vehicle is almost completely identified by its make, model, and year. Updates over an average 13-year life span are rare. (See Bureau of Transportation Statistics, 2014, for the distribution of vehicles by year.) In most countries (at least where there is left-hand drive), many aspects of driving are fairly consistent: the (a) input devices (steering wheel, brake, and accelerator), (b) method of operation, (c) location, and (d) primary displays (windshield and mirrors). In contrast, there is no consistency in the controls or displays for other vehicle interfaces. Furthermore, although new motor vehicle models are offered once per year, major changes typically occur once every four years or so. For nonautomotive products, computer software and hardware model upgrades occur almost continually. (However, many outside the industry do not know there are continual but minor changes throughout the year, known as rolling changes.) Thus, the hardware life cycles of the two contexts are quite different, and except for Tesla, motor vehicle software is almost never updated in a major way unless there is a product recall.

As of 2010, there were slightly more than 1 billion vehicles in operation worldwide excluding off-road vehicles and heavy construction equipment. About 1/3 of that 1 billion are commercial vehicles, primarily heavy trucks. There are more vehicles in the United States than anywhere else, followed by China, and then Western Europe. China is the largest vehicle market. However, relative to the United States and Europe, much less has been published about Chinese drivers or travel behavior.

Beyond the overall production and sales data, vehicle sizes and types sold vary among countries, and even regions within a country. People on the east and west coast of the United States think of cars as the primary means of personal transportation. Yet, in other parts of the United States, trucks, especially pickup trucks, predominate. Sport utility vehicles (SUV) are now a very important part of the global market but they did not exist until about 2002. In the United States, light truck and SUV sales often exceed that of passenger cars (Cars Move America, 2016). Table 40.2 shows the best selling vehicles in the United States. The Ford F-150 pickup truck has been the sales leader for many years.

One of the major unknowns for the future is the effect of ridesharing on vehicle type (Green, 2016). If ridesharing predominates, what types of vehicles will people want? To provide privacy in an otherwise shared space, will they want a private cabin (or minicabin the size of a toilet stall) or a completely open and shared space (as in a stagecoach)?

What are some of the Topics of Concern (and Solutions)?

Driver distraction is a topic that has received considerable attention in government reports (e.g., Ranney, 2008), has been a series in *The New York Times* (Driven to Distraction, 2010), and even has its own government web site (https://www.nhtsa.gov/risky-driving/distracted-driving). Although there are many potential distractions a driver might encounter, this chapter focuses only on those related to HMIs.

Table 40.2 United States motor vehicle sales for 2015 as of October.

Rank	*Vehicle*	*Type*	*October 2015 year to date*
1	Ford F-Series	pickup truck	629,951
2	Chevrolet Silverado	pickup truck	492,551
3	Dodge Ram	pickup truck	371,574
4	Toyota Camry	car	361,111
5	Toyota Corolla	car	306,693
6	Honda Accord	car	294,935
7	Honda CR-V	SUV	288,531
8	Nissan Altima	car	283,372
9	Honda Civic	car	277,538
10	Ford Escape	car	257,731
11	Toyota RAV4	SUV	256,178
12	Ford Fusion	car	255,143
13	Nissan Rogue	SUV	238,146
14	Chevrolet Equinox	SUV	236,128
15	Ford Explorer	SUV	210,894
16	Hyundai Elantra	car	209,830
17	Chevrolet Cruze	car	193,680
18	Ford Focus	car	180,287
19	GMC Sierra	pickup truck	180,174
20	Jeep Cherokee	SUV	178,785

Source: http://www.goodcarbadcar.net/2015/11/usa-october-2015-ytd-new-vehicle-sales-by-model.html.

Central to the issue of distraction is determining how it is to be identified. In brief, in order to drive, one must look at the road, and therefore the time the driver is not looking at the road is an indicator of distraction. A number of algorithms have been developed to predict the crash risk associated with distraction (Liang, Lee, & Yekhshatyan, 2012). These algorithms consider the number of glances associated with the task (the most important factor), the duration of each task, and the time between glances. Generally, crash risk is a power function of glance duration, where the exponent may be greater than 1.

In addition to these external influences, the auto industry is strongly influenced by customer feedback on its products, especially as expressed in the J. D. Power Initial Quality Survey (IQS, measured at 90 days of ownership) and the Vehicle Dependability Survey (VDS, after three years of ownership) (http://www.jdpower.com/cars/awards/Initial-Quality-Study-by-Segment/567ENG; U.S. Vehicle Dependability Study VDS, 2016). The details of J. D. Power surveys are not well known outside the auto industry because the reports are copyrighted and very costly. They are conducted carefully and the sample sizes are large. Nonetheless, there are sometimes halo effects, for example where an interface in a Lexus (made by Toyota) will receive a better rating than an identical interface in a Toyota. Users of the surveys are primarily interested in the relative ratings of their products to the competition, not the absolute product ratings. The questions relating to various aspects of driver interfaces receive considerable attention.

Cell phone problems (really smartphones)

Probably no single topic related to driving has received more attention than driver use of phones. This could be because smartphones are commonly used high-tech devices that appear to be distracting, or because there have been well known crashes associated with it, or for a variety of other reasons. Meta-analyses and other reviews show drivers taking longer to respond to brake lights of lead vehicles, departing from the lane more often, and exhibiting other undesired characteristics while using a phone (Caird, Scialfa, Ho, & Smiley, 2006; Collet, Guillot, & Petit, 2010a, b; Horrey & Wickens, 2006; McCartt, Hellinga, & Bratiman, 2006).

However, the most definitive indications of crash risk are odds ratios computed from real-world crash data. The estimates of crash risk associated with phones have fluctuated over time. The first widely recognized estimates of phone crash risk (used versus not used) were reported by Redelmeier and Tibshirani (1997) as 4.3 overall. Interestingly, hands-free units had a greater risk ratio (though not significantly) than hand-held units (5.9 : 1 versus 3.9 : 1). Other data (Koushki, Ali, & Al-Saleh, 1999; McEvoy et al., 2005; Violanti & Marshall, 1996) suggested similar risk ratios. In contrast, later data suggests a much lower risk (Dingus & Klauer, 2008; Highway Loss Data Institute, 2009; Farmer, Braitman, & Lund, 2010). However, in more recent research using the SHRP 2 data, Dingus et al. (2016) found the odds ratio for dialing using a hand-held phone to be 12.2 : 1 and for texting with a hand-held phone to be 6.1 : 1. See also Dingus et al. (2016); Dingus, Hanowski, and Klauer (2011), Klauer et al. (2014), and Young (2015). Smartphone issues are addressed in greater detail later in this chapter.

These changes reflect improvements in how the data are collected, which in turn lead to changes in the estimates. One key point is that it is not that phones are "bad" per se but the crash risk depends upon the task the driver is performing with the device, with tasks that are visually demanding, such as dialing phone numbers and texting, having greater risk than many other tasks. See Green, George, and Jacob (2003). What is particularly notable about texting is that the task duration (exposure) can be long, and the number of crashes depends upon exposure. What is unknown in the future is what tasks these devices will support and drivers will carry out. Of particular concern is broadcast video and video calling because they are visually demanding tasks. Curiously, one task of particular concern is answering the phone. Although the task duration is brief, people tend to respond immediately, sometimes independently of the traffic situation (Nowakowski, Friedman, & Green, 2001). Answering the phone should not usually be more important than driving but people behave otherwise out of habit. Personal observation suggests that most people, no matter who they are talking to, if they hear their phone ring, will at least check to see who is calling (almost immediately) and sometimes answer the phone, even if the caller is far less important that the person in front of them to whom they are speaking.

For additional information on distraction research see Bao, Flannagan, Xiong, and Sayer (2014); U.S. Department of Transportation (2017); Flannagan, Bao, and Klinich (2012), and Xiong, Bao, and Sayer (2014).

Problems with navigation systems

Entering a street address while driving using a visual-manual interface was probably the first task to raise concerns about driver distraction (e.g., CBC, 2010; Farber et al., 2000; Nowakowski, Utsui, & Green, 2000; Steinfeld, Manes, Green, & Hunter, 1996).

These concerns continue but, to a large degree, smartphones and other devices are supplanting in-vehicle navigation systems (Automakers trail in the great gizmo race, 2011). The major challenge is not looking at the map, but the destination designation task. Smartphone interfaces that allow for free form entry and have ready access to large databases often have short entry times. However, there are still issues when the address is not correct, is incomplete, or specifies a neighborhood in a city (e.g., Manhattan instead of New York City). There are also issues with similar street names but a different suffix (is it 1015 Peachtree, Peachtree Road, Avenue, Boulevard, Place, or something else?), and when map directions are part of the street name (4015 North Broad Street or 4015 South Broad Street) (Green and Park, 2013). Because of their limited databases, in-vehicle systems tend to have problems with finding points of interest (is Cobo Hall in Detroit a civic center, a community center, or in some other category?), which is not a problem for smartphone navigation systems.

Menu interface problems (especially for music selection)

Smartphones provide people with access to large music libraries. Libraries with several thousand songs are not unusual, and that number will surely increase in the future. The problem is that drivers try to retrieve songs, albums, and playlists while driving, possibly from the car itself, but often from a smartphone or off-board system or device. For in-vehicle devices, drivers may need to go through multiple menus to get to the desired selection. Furthermore, selections can be shuffled, transferred, renamed, and sorted, all while driving. For some research on this topic, see Bayly, Young, and Regan, 2008; Chisholm, Caird, and Lockhart, 2008; Garay-Vega et al., 2010. The solution seems to be free form entry such as that found in smartphones. Typically, this type of interface requires significant computational power that is currently only available off board.

Speech interfaces—are they the future?

There are some who believe that speech interfaces are the solution to the problem of information access while driving. There is good evidence to suggest that speech interfaces can be less distracting than visual-manual interfaces (Barón & Green, 2006; Garay-Vega et al., 2010; Maciej & Vollrath, 2009; Owens, McLaughlin, & Sudweeks, 2010; Shutko, Mayer, Laansoo, & Tijerina, 2009; Tsimhoni, Smith, & Green, 2004), although they are not without their problems (Chang, Lien, Lathrop, & Hees, 2009). (See also Lai, Cheng, Green, & Tsimhoni, 2001; Lo, 2013; Lo & Green, 2013; Mehler et al., 2015; Nass et al., 2005; Reimer, Mehler, Dobres, & Coughlin, 2013; Tsimhoni, Winter, & Grost, 2009; Wang, Winter, & Grost, 2015; Winter, Grost, & Tsimhoni, 2010; Winter, Tsimhoni, & Grost, 2011.) At this point, use of speech interfaces is uncommon because recognition performance is not very good. However, the considerable success of Ford Sync® (Shutko & Tijerina, 2011) and opportunities to process speech off board may change the situation. If anything, speech assistants such as Siri, Cortana, Google Now, and Alexa are showing promise (Chen, 2016; Kelly, 2015).

Workload managers—are they the future?

Given the concern for overload, one potential solution is to measure the primary task workload and then regulate the secondary tasks a driver can do at any moment using a workload manager (Green, 2000b, 2004; Hoedemaeker, de Ridder, and Janssen, 2002; Michon, 1993; Piechulla, Mayser, Gehrke, & König, 2003). As initially conceived, such systems would use data from four sources: (a) the navigation system (such as lane width and radius of curvature), to assess the demands due to road geometry, (b) the adaptive cruise control system (headway and range rate to vehicles ahead), to assess traffic demands, (c) the traction control system, to assess road surface friction, and (d) the wipers, lights, and clock, to assess visibility. This information—along with information on the driver (e.g., age) and the specific visual, auditory, cognitive, and motor demands of each in-vehicle task—could be used to schedule the occurrence of in-vehicle tasks. Thus, when driving on a curving road in heavy traffic in a downpour, incoming phone calls could be directed to an answering machine and the 30,000-mile maintenance reminder could be postponed. When the driving task demand is low, drivers could have access to a wide range of functions. Being able to predict the momentary workload of driving reliably, however, has proven to be very difficult.

A new potential application is for partially automated vehicles. If a situation arises where the vehicle needs to hand control back to the driver, an important question is how long that handover could potentially take and if the driver is prepared to accept the handover. Part of that decision involves a workload manager assessing the driving situation and determining if the driver could handle the situation as well as how much time is required for the driver to resume control.

What Measures and Statistics of Safety and Usability are of Interest? (The Empirical Measurement Issue)

As was noted earlier, one of the three principles essential to usability is empirical measurement. Superficial impressions suggest that the measurement of usability of office computer and Web applications, and the measurement of the usability of driver interfaces are quite similar. In an office, one measures task completion time, errors, and ratings of ease of use (Hornbaek & Law, 2007).

In a typical laboratory for examining driver interfaces, those same measures may be used to assess the usability of an in-vehicle or carry-in device task. However, other driving-specific measures, as listed in Table 40.3, may also be obtained, especially in driving simulators and on-the-road evaluations. (See Green, 1995a, b, c.) In the past, a major problem was that most researchers did not define how these measures were collected, making studies difficult to compare, and when they were defined, definitions were inconsistent. This problem is being resolved by using the definitions in Society of Automotive Engineers (SAE) Recommended Practice J2944 (Green, 2012, 2013; Society of Automotive Engineers, 2015). For example, does a lane departure mean a vehicle is about to depart, is departing, or has departed the lane? The definition that is most appropriate depends on how the measure is to be used, and therefore could refer to the inside, middle, or outside edge of pavement marking. These three alternatives differ by about a foot. In fact, SAE 2944 currently defines 11 different ways in which a lane departure could be identified. Similarly, headway can be measured to the front or rear of a lead vehicle, a difference of about 15 to 16 feet for a car and 55 feet for a tractor trailer—all substantial differences. In fact, the proper name for the distance between the front of a vehicle and the rear of

Table 40.3 Some examples driving-specific usability measures and statistics.

Category	*Statistic*	*Options (Variations) in J2944*
Lateral	steering wheel angle (mean, standard deviation)	
	steering wheel reversals (number, reversal rate)	amplitude method, amplitude and velocity thresholds method
	lateral position (mean, standard deviation)	relative to lane center, relative to mean path driven, relative to lane edge
	lane departures (number, duration, magnitude, etc.)	11 methods, see SAE J2944
	time to line crossing (mean, minimum, inverse)	acceleration method, velocity method, trigonometric
	steering entropy	1999 Boer method, 2005 Boer method
Longitudinal	number of collisions (number, mean impact velocity)	
	time to collision (minimum, inverse, minimum adjusted, time exposed, time integrated, etc.)	acceleration method, velocity method
	gap (mean and standard deviation of time or distance)	
	required deceleration	
	speed (mean, standard deviation)	
	speed drop during a task (mean)	
	braking events greater than some g threshold (number)	
Visual	glances (number, mean duration, number > 2 seconds, maximum)	
	percentage of off-road glances greater than 2 s	
	total eyes-off-the-road time	

Note: For definitions of some of these measures, see SAE J2499 (Society of Automotive Engineers, 2010a).

a vehicle ahead is gap. Those doing research on driving performance are strongly encouraged to use the terms in SAE J2944 and should expect in the future that conferences and journals will reject papers not using J2944 specified terms. This does not mean that other terms cannot be used—only that they need to be defined with the level of specificity in SAE J2944.

In addition to the standard measures of driving performance, a host of other performance measures are often collected in driving studies: these include (a) ratings of workload (NASA TLX), (b) measures of situation awareness, and (c) measures of object and event detection (pertaining to traffic). (See Johansson et al., 2004; Roskarn et al., 2002; Tijerina, Angell, Austria, Tan, &Kochhar, 2003.) In fact, simulator and on-road studies of telematics typically involve anywhere from 10–30 dependent measures, although in operational field tests the collection of several hundred measures in real time is common (General Motors Corporation, 2005; LeBlanc et al., 2006). Of them, TLX seems to be most popular, but TLX does not seem to provide much insight, and comparisons between studies are rarely made, in part because there are no anchors to support them. To provide somewhat more focus, some researchers use the Driving Activity Loading Index (DALI), a TLX-like derivative (Pauzie, 2008).

A major challenge in assessing the safety and usability of driver interface is dealing with the tradeoffs that drivers naturally make. The impression is that when preoccupied with an in-vehicle task, drivers lose awareness of the driving context—that is, situation awareness. Drivers attempt to compensate by slowing down (to make driving easier), allowing for larger gaps and, if very preoccupied, paying less attention to steering (so lane variance and the number of lane departures increase). However, drivers can respond in strange ways. For example, if asked to use two different in-vehicle systems, one of which is not well designed, they might attempt to maintain equal performance on both: slow down more for the more difficult interface, but compensate by having better steering performance for the poorer interface. Assessment is difficult because the tradeoff functions for all of these measures are unknown. One strategy used to overcome the tradeoff problem is to minimize the opportunity for tradeoffs. For example, this could include using cruise control to fix the speed (and in some cases, headway) and provide incentives and feedback to maintain driving performance, so only task time and errors would trade off. Of course, those constraints change the nature of the driving task.

How are Driver Interfaces Evaluated?

The classical usability lab has (a) a one-way mirror, (b) multiple cameras, (c) video editing equipment, (d) audio mixers, and (e) at least two rooms, one for the subject and one for an experimenter. Currently, Morae (Techsmith.com) is the most popular software application for mixing the audio and video, recording the screen, recording keystrokes and mouse actions, and logging comments.

For driving research, there is no single favored platform for driving simulators, the driving equivalent of a usability laboratory. In the United States, the most popular driving simulator vendors/simulators are NADS Minisim (University of Iowa), Realtime (Realtime Technologies, Inc.), STISIM (Systems Technology, Inc.), and Drive Safety, Inc. The author has used the OpenDS (OpenDS.eu) platform for instructional purposes because there is a useful free version and the software can be readily modified because source code is provided. Others have used it for research as well. There is also interest in using Unity-based software for developing driving simulators as it is a gaming platform. The future is uncertain. Will some of the current vendors disappear because their platforms use old technology? Will some new technology come along that supplants what is state of the art now? What is important to bear in mind is that developing a library of test scenarios can take years of work, whereas the hardware becomes obsolete very quickly.

The major challenge in using driving simulators is the lack of standard scenarios and the significant time required to develop scenarios of any type, which is why researchers are reluctant to change platforms. Although image resolution has improved over time both for large screen monitors and projection displays, and those improvements will continue, the most significant developments over the next few years will be the wider use of motion systems from driving games and virtual reality displays. These developments will enhance the reality of the simulator experience and eventually reduce motion sickness. However, most systems still cannot display images resolutions consistent with the resolution of the human eye. This is most noteworthy for navigation studies, where highway signs need to be oversize so they can be read at the intended distance ahead. Displays with 4k and 8k resolution can help solve that problem.

Similar to the case for driving simulator, in the United States there is no favored vehicle instrumentation platform, with each organization creating its own. Typically, some sort of laptop CANalyzer software / hardware combination is used to collect data (speed, steering wheel angle, gaps to other vehicles, etc.) from the Controller Area Network (CAN) bus in real time. One particularly interesting innovation is the use of smartphones to collect basic data (Johnson, and Trivedi, 2011). These systems greatly simplify instrumenting a vehicle of interest. The major challenge is that the CAN codes are proprietary and secret, so vehicle manufacturer cooperation is essential.

Complementing and synched with the CANalyzer are several video cameras, either using web cameras or higher quality cameras. Typical camera views include the forward scene, the driver's face, and the cab interior. Increasingly, the left and right side views and the rear are also recorded. The camera images help make the digital data collected by the CANalyzer easier for the analyst to understand and greatly aid those who are trying to understand what the analyst found. Files can be very large but storage space is becoming a secondary issue. Unfortunately, at the current time, most commercial statistical packages do not allow for integration of video data or provide tools for its analysis.

A major change in recent years has been the conduct of naturalistic driving studies: ACAS (Ervin et al., 2005); 100-car study (Dingus et al., 2006), Safety Pilot, Strategic Highway Research Program (SHRP 2, Campbell, 2012), etc. In these multimillion-dollar studies, often funded by a national government department of transportation, subjects are given a car, typically for a month, to use as their own vehicle. Usually, subjects can drive wherever and whenever they want. There is no experimenter in the vehicle. The test vehicles have a complete set of instrumentation, including cameras that are recording everything while the car is being driven, CANalyzer output, GPS to determine location, and most importantly, a digital cellular connection so the data can be uploaded remotely. In some cases, the test vehicle may be the subject's own vehicle.

In a typical experiment, the subject drives the vehicle for about a week to collect baseline data and then some system of interest, for example, adaptive cruise control, is enabled for subjects to use as they see fit. Although one can learn a great deal from driving simulator and accompanied experimenter studies, for technology that could change driving in substantial ways, one needs the data from naturalistic studies to understand real use before mass installation of some new technology in production vehicles. A major challenge has been making the data publically available for analysis, which has notably occurred for the SHRP 2 project (Transportation Research Board, n.d.). The challenge is that the most important data are the subject's expressions, where they look, and sometimes what they say. With this data in hand, the subject may be identifiable, raising privacy concerns.

Simulator and accompanied experimenter studies have benefited from improvements in eye gaze recording systems. They are particularly important because where drivers are looking helps analysts determine what they are doing. Furthermore, if glances away from the road are frequent and extended, safety is compromised. The latest trend is toward finding ways to obtain useful data from very low-cost eye-gaze systems (hundreds of dollars of hardware). The author has found that some useful data (for drivers without glasses) can be obtained using a Gazepoint system (gazept.com). Currently, the system can only handle one screen, a limitation likely to be overcome in the future. Smarteye (Smarteye.se) and SeeingMachines (seeingmachines.se) are among the two more popular vendors of high-quality systems. Those seeking additional information on this topic should examine Chapter 21.

A major thread of HMI research is towards understanding how drivers deal with automation. The conventional approach is to create real code for the automation and see how drivers respond to it, either in test vehicles or in a driving simulator. This approach is time consuming and costly. The popular alternative is the Wizard of Oz method, in which a hidden experimenter (the wizard) simulates the actions of the automation. So, for a driving simulator, there are two steering wheels and two sets of foot pedals. When the subject presses a button for the automation to drive, the hidden experimenter drives, returning control to the subject when either the subject requests it by pushing a button or when the wizard pressed a button to return control to the subject (as if the automation failed). Similarly, to simulate a self-driving left-hand-drive car, a right-hand-drive car is used, with a partition between the subject in the left front seat and the hidden experimenter in the right front seat (Baltodano, Sibi, Martelaro, Gowda, & Ju 2015; Green, 2016). In most cases, subjects have no idea the automation is simulated, and even if they do, if the task is sufficiently engaging, then they perform as if the simulation was real.

A complete assessment of a driver interface often uses of a variety of methods – standalone simulations, driving simulator evaluations, and on-the-road evaluations— as each method has its own strengths and weaknesses and is most appropriate for difference questions that arise at different phases of the product development (Table 40.4). (See Green, 1995a, for additional details.)

What Design Documents Exist for Driver Interfaces?

What types of documents are there?

Although feedback from the empirical measurements just described is important, motor vehicle design is strongly influenced by industry and government regulations, and international standards (Jeong & Green, 2013). For office applications, compliance with design guidelines, commonly called style guides, is achieved by providing application program interfaces (APIs) in the operating system that assure that widgets such as windows, menus, and so forth, all work in a consistent way. For driver interfaces, similar, publicly available, product-specific interface guidelines do not exist, but there are other types of important written materials (Green, 2001a, 2001b; Schindhelm et al., 2004).

Motor vehicle standardization activities occur in the SAE Safety and Human Factors Steering Committee and the International Standards Organization (ISO) Technical Committee 22 (Road Vehicles), Subcommittee 39 (Ergonomics), Working Group 8 (Telematics). In addition, activities relating to automated vehicles are occurring within the SAE On-Road Autonomous Vehicles (ORAV) Standards Committee and the ISO Technical Committee 204 (Intelligent Transportation Systems), although interface issues should be the attention of the prior SAE and ISO groups.

Table 40.4 Evaluation contexts.

Evaluation	*Method*	*Comments*
Focus groups	Groups of 8–12 people demographically similar to customers sit around a table and discuss a product or service guided by a facilitator Camera is often behind one-way mirror Generally done in multiple cities (one or two groups/city) Often conducted by a marketing firm	Useful in getting ideas for product concepts, but not predictions of the safety or usability of new products because the products have not been used Approach is sometimes used by manufacturers when a usability test might be more appropriate Generally, no quantitative data Essential to report actual quotes from participants, not what the facilitator recalls
Clinics	Customers in various cities are given the opportunity to experience a new product and its competition, often two or three vehicles, side by side Customers say which product or feature they prefer Performance data often not collected	Only exposes users to a limited number of options Approach is commonly used by industry Because the results are highly proprietary, published studies are rare
Part task simulation	Sample of users operate the device (e.g., computer-simulation of a new radio, occlusion test) and user task times, errors, and comments are recorded Test facility is not sophisticated	Not done that often Used during initial stages to assess tasks Use of occlusion method is very common Less costly that simulator and on-road evaluations
Driving simulator	Typically driving simulators are fixed base (no motion) and cost $25,000 to $250,000 each, but the major cost is for a full-time engineer (or engineers) or several graduate students to operate the simulator Simulators at manufacturers tend to be in 1–3 million dollar range, though some are much more (e.g., Ford is about $10,000,000). 1–5 projectors with a total 40–210 degree forward field of view, real vehicle cab, steering system with torque feedback, and realistic sound rear image may be projected or mirrors may be replaced with small LCDs Facility can require considerable space (e.g., 1,000 square feet) Generally requires large number of fixed small (lipstick or smaller) cameras	Operation requires considerable experience Simulator sickness is a major problem, especially for wide field of view and older drivers Each experiment requires construction of a test road/world and scripting the behavior of vehicles and pedestrians Best-known vendors in the United States are University of Iowa (NADS MiniSim), Systems Technology Inc., Realtime Technologies, DriveSafety and OKTAL in Europe. OpenDS has a free option.

(*Continued*)

Table 40.4 (Continued)

Evaluation	*Method*	*Comments*
Instrumented vehicle on test track or public roads	Production vehicle is fitted with cameras aimed at driver, forward scene, instrument panel, and lane markings, and with sensors for steering wheel angle, brake pressure, speed, and headway Eye fixation system may also be provided System of interest is also installed For lower cost systems, driving performance data is collected using a CANalayzer. Otherwise, the data collection system is custom	When the subject's own vehicle is used, only high volume vehicles are used to reduce the amount of customization and installation time required, and avoid damaging the subject's vehicle.
Operational field test	Compact instrumentation is installed in a fleet of vehicles (10–50) Each vehicle is borrowed by a potential user for a week, a month, or even a year—a month is most common Driving performance is surreptitiously recorded by the vehicle Unlike an instrumented car, continuous video is not recorded In addition to data recorded by the instrumented vehicle, GPS-determined location is also recorded Vehicles are periodically polled for data (and data is automatically dumped) by an independent digital cellular phone Test is confined to a single metropolitan area but tests in multiple areas are being planned	Each test requires unique instrumentation Tests are very expensive ($10,000,000 to $40,000,000) and can only be conducted with significant government support Experiment generally lasts several years Planning stage for experiments takes several years At any given time, there may only one operational field test in progress in the United States.

Motor vehicle design documents fall into five general classes: (a) principles, (b) information reports, (c) guidelines, (d) recommended practices, and (e) standards. Principles give high-level recommendations for design and are similar to those found in office HCI applications, such as "design interfaces to minimize learning."

"Information report" is a term used by the SAE to refer to a compilation of engineering reference data or educational material useful to the technical community. Information reports do not specify how something should be designed, but provide useful background information.

Guidelines give much more specific advice about how to design an interface element. For example, Guideline 9 in Chapter 7 of Green, Levison, Paelke, and Serafin

(1993) stated, "Turn displays should show two turns in a row when the turns are in close proximity" (p. 41), where close proximity means "0.1 miles apart or less." The impact of guidelines can depend on the issuing organization. For example, automotive design guidelines written by research organizations have no real authority. Guidelines written by the International Organization for Standardization (ISO), although technically voluntary, can become requirements, because in some countries, type approval (approval for sale) requires compliance with ISO guidelines. For vehicle models sold worldwide, global manufacturers find building common vehicle systems that comply with ISO standards to be less costly than building noncompliant, country-specific systems. In Japan, the Japan Automobile Manufacturers Association (Japan Automobile Manufacturers Association, 2004) has a set of guidelines for navigation systems. Although theoretically voluntary, "requests" from the National Police Agency make the JAMA guidelines a requirement for all original equipment manufacturers (OEMs) in Japan. One could view the U.S. DOT driver distraction guidelines in a similar manner.

"Recommended practice," a term used by the SAE, refers to specifications for a material, product design, or design or test procedure that are intended to guide standard engineering practice, often because they have not have gained broad engineering acceptance. Commonly, a recommended practice is followed. A product liability action against a product (especially in the United States) is extraordinarily difficult to defend if the product design and evaluation deviate from recommended practice.

In some sense, a standard is a recommended practice that is a broadly accepted engineering practice, and must be followed. In the SAE context, "must" has an unusual connotation because an SAE standard is technically voluntary as the SAE has no enforcement powers. However, in a product liability context, a product not complying with an SAE standard is almost not defendable, and is unlikely to be purchased from a supplier by a motor vehicle manufacturer.

Non-ISO documents

Table 40.5 provides a summary of the design document activities to date excluding those of the ISO (described later). As indicated in the table, the EU guidelines are quite brief and are merely statements of very general principles (for example that interfaces should be simple to operate), though they serve as the basis of the Alliance guidelines described later. Documents that are followed are JAMA (as was described previously), the Alliance guidelines (because the Alliance members agreed to follow them, possibly to avoid government regulation), and SAE and ISO guidelines (both of which are accepted industry practice and are described in the sections that follow). For all of these guidelines, see http://www.umich.edu/~driving/guidelines/guidelines.html for unofficial electronic copies. (To avoid copyright problems, only draft versions have been posted for SAE and ISO documents.) Readers are cautioned that all of these documents are updated every few years, and they should verify that they are using the most recent version with the authoring organization. Although these documents have required a significant development effort, the number of times they are cited in the research literature is limited.

Table 40.5 Major non-ISO Driver Interface Guidelines and Recommended Practices.

Common name	*Reference*	*Size (pages)*	*Comments*
Alliance guidelines	Alliance of Automobile Manufacturers (AAM) (2006); version 3	90	Elaboration of the EU principles with details on the method and rationale, used by almost all manufacturers in the United States; key sections are principles 2.1 and 2.2, which still need development
Battelle guidelines	Campbell, Carney, and Kantowitz, (1997)	261	Voluminous document with references to interface design, heavy on trucks. User interface has been said to have a Windows OS flavor, includes physical ergonomics information (e.g., legibility, control sizes) which are not included in the UMTRI guidelines.
EU guidelines	Commission of the European Communities (2008)	42	Mostly vague platitudes
HARDIE guidelines	Ross et al. (1996)	480	Early set of European guidelines, less data than UMTRI or Battelle
JAMA guidelines	Japan Automobile Manufacturers Association (2004)	15	First set of detailed design guidelines for driver interfaces. These guidelines are voluntary in Japan but followed by all OEMs there and sometimes by aftermarket suppliers. Some aspects are particular to Japan. Device location restrictions are important.
SAE J2364 ("15-second rule")	Society of Automotive Engineers (2004a)	13	Specifies the maximum allowable task time and test procedures for navigation system tasks performed while driving for systems with visual displays and manual controls; also describes an interrupted vision (visual occlusion) method as well; See also SAE J2365.
SAE J2365 (SAE calculations)	Society of Automotive Engineers (2002)	23	Method to compute total task time for tasks not involving voice, used early in design to estimate compliance with J2364
TRL guidelines	Stevens, Quimby, Board, Kersloot, & Burns (2004)	56	Expansion of simple check list
UMTRI guidelines	Green, Levison, Paelke, and Serafin (1993)	111	First set of comprehensive design guidelines for the United States. Includes principles, general guidelines, and specific design criteria with an emphasis on navigation interfaces
Driver-Vehicle Interface (DVI) guidelines	See Campbell, Doerzaph, Richard, and Bacon (2014) for an overview	to be determined	Development is complete or is about to be completed

ISO Documents

Much of the ISO activity in recent years has occurred under the auspices of International Organization for Standardization Technical Committee 22, Subcommittee 39 (ISO TC 22/SC 39—Ergonomics Applicable to Road Vehicles, in particular, Working Group 8 (WG8—Transport, Information, and Control Systems or TICS); Green, 2000a). (Note: formerly, the ergonomics subcommittee was number 13.) WG8 has about 35 delegates from the major vehicle-producing nations, with the most active members, approximately 15, appearing at meetings held two to three times per year, usually in Europe.

Table 40.6 shows the standards and technical reports developed (or in progress) by Working Group 8 that pertain to driver interfaces. For the complete list, go to the ISO TC 22/SC 12 portion of the ISO website (www.iso.org). For a variety of reasons, ISO documents emphasize measurement methods and organization over specifications and safety limits. To promote international harmonization, national standards organizations, technical societies (e.g., SAE), and government organizations (e.g., U.S. Department of Transportation) often permit ISO standards to supersede their own standards, so ISO standards are very important.

Note: ISO documents follow a very well defined, three-year process through several stages (Preliminary Work Item [PWI], Committee Draft [CD], Draft International Standard [DIS], Final Draft International Standard [FDIS], and International Standard [IS]) as they are passed from the working group to the subcommittee to the technical committee, and finally to the secretariat for review and approval. The major hurdles are the working group and subcommittee, where passage requires approval by two-thirds of the participating nations. The emphasis of this process is on building a voluntary consensus. Some items that are more informational in nature become technical reports instead of standards. Because of the limited number of experts available, WG8 is very selective in adding items to its work program.

For additional information on ISO standards, readers should examine Chapter 3, Volume 1 of this book on that topic.

U.S. DOT (NHTSA) Guidelines

An important development since the last edition of this book has been the release of the distraction guidelines (U.S. Department of Transportation, 2013) by the National Highway Traffic Safety Administration, an organization within the U.S. Department of Transportation. These guidelines are not to be confused with the Driver-Vehicle Interface Guidelines (Campbell, Doerzaph, Richard, & Bacon, 2014). Although only a guideline, it appears that all manufacturers and suppliers want to produce interfaces that comply with them because of the consequences of failing to comply in a product liability lawsuit. In fact, the degree of attention and compliance with them is far greater than any other document mentioned in this section.

The guidelines contain recommendations that limit what a driver should be able to do in a moving vehicle. There are a number of tasks that are not permitted while driving, such as manual entry for text messaging, watching entertainment videos, participation in video phone calls, and automatically scrolling lists, all tasks that are visually intensive.

Table 40.6 ISO TC 22/SC 39/WG 8 published documents.

Status	*Topic*	*Summary*
Std 2575	Symbols	Contains symbols for controls and display that must be used (upper beam, hazard, wiper, airbag malfunction, cruise control, lane departure, etc.)
Std 3409	Lateral spacing of foot controls	For cars, spacing of accelerator, brake, and clutch pedals
Std 3958	Hand-control reach	Provides a reach envelope for different proportions of men and women, based on U.S. data for three-point restraint and lap belt only
Std 4040	Location of controls and displays	Very few real constraints
TR 12204	Critical warnings	General information on safety and time critical warnings including hazard perception and visibility
Std 12214	Direction of motion stereotypes	Provides direction in which controls should move to be consistent with user preferences (e.g., clockwise movement of a control or display for increase)
TS 14198	Calibration task	Describes two secondary tasks (critical tracking task (CTT), surrogate reference task (SURT)) that can be used to calibrate the demands of other secondary tasks. SURT is a search task that involves find one circle that is slightly larger than others in an array of circles. CTT is a tracking task that becomes increasingly difficult.
Std 15005	Dialogue management principles and compliance procedures	Provides high-level ergonomic principles (compatibility with driving, consistency, simplicity, error tolerance, etc.) to be applied in the design of dialogues that take place between the driver of a road vehicle and an in-vehicle information system while the vehicle is in motion. Provides general directions on how to test for compliance.
Std 15006	Specifications and compliance procedures for in-vehicle auditory presentation	Provides requirements for auditory messages including signal levels, appropriateness, coding, etc., along with compliance test procedures
Std 15007, part 1 and TS part 2	Measurement of driver visual behavior	Generally describes video-based equipment (cameras, recording procedures, etc.) and procedures (subject descriptions, experiment design parameters, tasks, performance measures, etc.) used to measure driver visual behavior
Std 15008	Legibility (visual presentation of information)	Provides requirements for character size, contrast, luminance, etc., and specifies how they are to be measured
Std 16121	Bus workplaces	Basic design requirements, visibility, controls and displays, cabin environment (four parts)
Std 16352	Warning system messages	Literature review (circa 2002) of warning systems covers topics such alarm theories; the design of visual, auditory, and tactile warnings; redundancy; etc. The report contains summaries of a significant number of studies.

(*Continued*)

Table 40.6 (Continued)

Status	*Topic*	*Summary*
Std 16673	Visual occlusion method	Describes how to conduct a visual occlusion evaluation of the demand of in-vehicle tasks
TS 16951	Message priority	Provides two methods for determining a priority index for in-vehicle messages (e.g., navigation turn instruction, collision warning, low oil) presented to drivers while driving. For one method, priority is based on criticality (likelihood of injury if the event occurs and urgency (required response time)), determined on four-point scales by experts.
Std 17287	Suitability of TICS while driving	Generally describes a process for assessing whether a specific in-vehicle task, or a combination of tasks is suitable for use while driving
Std 16673	Occlusion method to assess distraction	Describes how the visual demand of a display can be assessed by periodically blocking (occluding) the driver's view of the display. Includes requirements for number and training of subjects, test hardware, viewing and occluded periods, and two metrics for data analysis.
Std 20176	H-point machine	Describes procedure for using version 2 of the H-point machine to locate human anatomical reference points in a seat
Std 26022	Lane change test	Describes procedure for testing the demand of in-vehicle tasks using a PC-based driving simulator. Subjects perform a number of lane changes, some of which occur while performing the task

Note: ISO 17488 (the decision response-time task or DRT) should be approved soon. In that task, subjects are present with either a peripheral light of a vibrating tactor on their neck. The mean response time and number of missed responses are indications of distraction.

In addition, two procedures are provided to assess other tasks, one involving visual occlusion and a second involving car following in a driving simulator. In the occlusion protocol, an age-stratified sample of 24 subjects (Table 40.7) performs the task of interest while wearing occlusion goggles (Figure 40.1) whose operation is described in the section that follows. The lower age limit (18) was set due to concerns about testing minors. Organizations may set an upper age limit (e.g., 65) if they have concerns about testing elderly subjects. Although motion sickness is more likely in older subjects and is an issue for simulator testing, older subjects (e.g., >65) are most likely to have problems with complex user interfaces and are essential to include in evaluations. Quite frankly, many people in this age category are readily available, licensed drivers, and if they are retired, looking for interesting experiences to serve as subjects.

In the occlusion method, while performing the in-vehicle task, the goggles cycle between being open (so subjects can see the driver interface) and closed (so they cannot, representing a glance to the road). The open and close times are 1.5 s each, and the subject has a maximum of 8 cycles (8 1.5 open and closed intervals, corresponding to a 12.0 s open time) to complete the task. Otherwise, the task should not be performed while driving.

Table 40.7 Subject sample for U.S. DOT distraction tests.

Age range	*Women*	*Men*
18–24	3	3
25–39	3	3
40–54	3	3
>55	3	3
Total	**12**	**12**

Figure 40.1 Visual occlusion goggles; open and closed states are shown. *Source:* http://www.translucent.ca/wprs/wp-content/uploads/lensanimation.gif

In the driving-simulator experiment, subjects follow a lead vehicle with a U.S. DOT prescribed speed profile. The task is considered acceptable if (a) 21 of the 24 subjects' mean off-road glance duration is less than or equal to 2 s, (b) 85% of the off-road glances made by 21 subjects during the data trial are less than 2 s, and (c) the sum of off-road glances is less than or equal to 12 s for 21 subjects. Obviously, to conduct this evaluation, one needs (a) a driving simulator that meets the U.S. DOT guidelines and (b) a reliable eye-gaze system that works with glasses and contacts, which are commonly worn by older subjects in the DOT-prescribed sample.

There is a great deal more to the test procedure than is described here, and those interested in using these evaluation methods should read the guidelines. For research using evaluation methods related to the guidelines, see Boyle et al., 2013; Dopart et al., 2013; Kidd, Dobres, Reagan, Mehler, and Reimer, 2015, and Ranney et al., 2011. The method is not without its critics (Kujala, Lesch, & Makela, 2014; Pournami, Large, Burnett, & Harvey, 2014).

SAE J2364 (the 15-second rule)

Other than the NHTSA distraction guidelines, few documents have generated as much discussion as SAE Recommended Practice J2364, sometimes referred to as the "15-second total task time rule" but more commonly as the "15-second rule." SAE J2364 establishes two procedures for determining if a task involving driver use of visual displays and manual controls is excessive while driving (Green, 1999c; Society of Automotive Engineers, 2004a, b). The practice does not apply to voice interfaces. Recently, the standard was frozen recognizing the importance of other design standards and guidelines, in particular, the NHTSA guidelines. Nonetheless, the criteria are still reasonable.

In brief, the SAE J2364 test procedure (Society of Automotive Engineers, 2004a) requires that 10 subjects between the ages 45 and 65 be tested. Each subject completes five practice trials and three test trials in a parked vehicle, simulator, or laboratory mockup. The test cases to be examined (e.g., addresses for destination entry) are to be representative of what is planned for production. Interestingly, the choice of the address can have a marked impact on the task time.

In the static method, the subject performs the task, with the duration being from when the subject is told to start until the goal is achieved. The interface complies with J2364 if the mean of the log of the task times is less than the log of 15 s. Logs were used to reduce the influence of long outliers. See SAE J2364 for additional details on the test method.

Static method times are reasonable estimates of the actual task times when parked. Times while engaged in real-world driving (where drivers alternate between looking inside the vehicle and looking at the road, evaluations that are much more expensive to conduct) can be estimated by multiplying static times by values of 1.3 to 1.5, depending on the workload of the primary task. The J2364 test procedure does not suggest that drivers can safely look away from the road continuously for 15 s.

Some have argued that use of static task time fails to identify interfaces requiring long glance durations. However, analysis of real products shows the primary risk is from tasks that take too long to complete (Young, 2012). In fact, it is very difficult to think of driver tasks for navigation systems that have short total task times but very long glance durations. In real driving, people truncate glances to the interior when the glances become too long but tend to complete tasks, even if they are unacceptably long. In practice, eliminating tasks with long completion times (the worst tasks) also eliminates many of the tasks with long glance durations.

Nonetheless, J2364 provides an alternative method involving visual occlusion. In that method, either the subjects wear LCD goggles described earlier, or vision to the device is otherwise periodically interrupted, simulating looking back and forth between the road and the device. Unlike driving, though, subjects do nothing in the occlusion interval. The device is visible for 1.5 s and occluded for 1 to 2 s, with 1.5 s being recommended. Compliance is achieved if the sum of the log of the viewing times is less than the log of 20 s. The essence of the procedure is similar to that of the NHTSA guidelines but the allowable task time is much longer.

Alliance Principles

The Alliance of Automobile Manufacturers (AAM), the trade association of 12 major manufacturers of automobiles in the United States (GM, Ford, Toyota, Mercedes, etc.), has devoted considerable effort to developing design guidelines/principles

(Alliance of Automobile Manufacturers, 2006). The Alliance guidelines are a detailed elaboration of the 24 principles in the EU guidelines, formally known as the European Statement of Principles on HMI but often referred to as ESOP 2 (Commission of the European Communities, 2008). As an example, Principle 1.1 of the Alliance Guidelines states, "The system should be located…in accordance with relevant… standards…" "No part…should obstruct any vehicle controls or displays…" (Alliance of Automobile Manufacturers, 2006, p. 13). Those guidelines seem obvious at a high level but defining precisely how they can be met can be quite difficult. Each principle has four parts: (a) rationale (usually quite detailed), (b) criterion / criteria, (c) verification procedures, and (d) examples. The release of the NHTSA guidelines has decreased the use and importance of the AAM principles.

The most important principle is 2.1: "Systems with visual displays should be designed such that the driver can complete the desired task with sequential glances that are brief enough not to adversely affect driving" (p. 39). Two alternative sets of criteria are offered. Alternative A states that "single-glance durations generally should not exceed 2 seconds," and task completion "should not require more than 20 seconds of total glance time." (Notice the use of the words "should," not shall.) There is debate as to what the percentile criterion for a single glance and the maximum task time should be (Go, Morton, Famewo, & Angel, 2006). Verification can be achieved by a visual occlusion procedure (1.5 s viewing time, 1.0 s occlusion time), or by monitoring eye glances directly using either a camera aimed at the face or an eye gaze monitoring system in either a divided-attention or on-road test. Note that the occlusion time is less than that in the NHTSA guidelines and SAE J2364.

Alternative B requires that the number of lane departures "should" not exceed the number associated with a reference task such as manual radio tuning, and that cars following headway "should" not degrade under those conditions, either. The radio tuning task also has served as the baseline task for the NHTSA guidelines. The verification procedure is stipulated to be driving on a divided road (either real or simulated) at 45 mph or less in daylight, on dry pavement, with low to moderate traffic. Additional details are provided describing the location of the radio, the stations to choose among, what constitutes a trial, subject selection (equal numbers of men and women between the ages of 45 and 65), and so forth. This test procedure bears some similarities to those in the NHTSA guidelines.

Although both procedures seem well described, additional details and constraints are needed to make those procedures repeatable. For example, the differences in performance between driving in "low" and "medium" traffic could be quite considerable and need to be quantified, as in Schweitzer and Green (2007). Criteria for acceptable levels of variation in speed and lane position are needed, criteria that could be developed from the data in Lai's (2005) dissertation or from Jamson, Wardman, Batley, and Carsten (2008). For many tasks that involve database searches (of address lists, song files, etc.), compliance with the principle will depend on the size of the database and the subject's familiarity with it.

A disadvantage of many of these procedures, especially those conducted on the road or requiring driving simulators with eye-gaze systems, is that they are extremely expensive and time consuming and can occur too late in the development process to have a useful impact. Therefore, the author strongly prefers simpler evaluation procedures such as J2364, and the task time estimates determined using SAE J2365 and Pettitt's calculation procedures described in the next section. At the present time, the auto manufacturers seem to prefer the test procedures in the NHTSA guidelines, in particular the occlusion procedure.

What Calculation Methods Can Aid Telematics Design?

SAE J2365 Task Time Calculations

SAE J2365 (Green, 1999a) was developed to allow designers and engineers to calculate completion times early in design, when the design is still a concept that can easily be modified. As with J2364, J2365 is for in-vehicle tasks involving visual displays and manual controls evaluated statically—that is, while parked (or in a benchtop simulation). SAE J2365 applies to both original equipment manufacturer and aftermarket equipment.

The calculation method is based on the Goals, Operators, Methods, and Selection rules (GOMS) model described by Card, Moran, and Newell (1980) and is discussed in Chapter 8 of Volume 1. The keystroke data was drawn from UMTRI studies of the Siemens Ali-Scout navigation system (Manes, Green, & Hunter, 1998; Steinfeld et al., 1996). Search times were based on Olson and Nilsen (1987–1988), and the mental time estimates were drawn from the Keystroke-Level Model (Card, Moran, & Newell, 1983) and UMTRI Ali-Scout studies. Thus, the times shown in Table 40.8 have been tailored for the automotive context. (See also Nowakowski et al., 2000.)

The basic approach involves a top-down, successive decomposition of a task. The analyst divides the task into logical steps. For each step, the analyst identifies the human and device task elements. Sometimes analysts get stuck using this approach because they are not sure how to divide a task into steps. In those cases, using a bottom-up approach may overcome such roadblocks. For each goal, the analyst identifies the method used. The analyst is advised to document each method using paragraph descriptions and then convert those descriptions into pseudo code (Figure 40.2). All steps are assumed to occur in series; multiple tasks cannot be completed at the same time.

Next, the pseudo code task description is entered into an Excel spreadsheet. The analyst looks up the associated time for each element listed in the table and sums them to determine total task time. By way of explanation, if a young user were to press a cursor key three times and no prior mental activities were required to decide to do so, the estimated time would be $C1 + C2 + C2 = 0.80 + 0.40 + 0.40$ or 1.6 s. To assist in understanding the process, the practice has provided a step-by-step example of entering a street address into a PathMaster/NeverLost navigation system, a popular U.S. product. For background on the calculation method, see Green (1999b). Those intending to use this method should examine the SAE document in detail as the space available here is not adequate to include all important details.

The J2365 approach makes a number of assumptions, many of which are also shared with the basic GOMS model. The model assumes error-free performance, which, while not likely, can be adjusted for by increasing the computed value by a percentage, often by 25%. Further, activities are assumed to be routine cognitive tasks, with users knowing each step and executing them in a serial manner. Again, adjustments in computed time can account for users sometimes forgetting what is next.

Table 40.8 SAE J2365 element times (s).

Code	Name	Description	*Time (s)* Young drivers (18–30)	Older drivers (55–60)
Rn	Reach near	From steering wheel to other parts of the wheel, stalks, or pods	0.31	0.53
Rf	Reach far	From steering wheel to center console	0.45	0.77
C1	Cursor once	Press a cursor key once	0.80	1.36
C2	Cursor ≥ twice	Per keystroke for the second and each successive keystroke	0.40	0.68
N1	Number once	Press the letter or space key once	0.90	1.53
N2	Number ≥ twice	Per keystroke for the second and each successive keystroke	0.45	0.77
L1	Letter or space 1	Press a letter or space key once	1.00	1.70
L2	Letter or space ≥ twice	Per keystroke for the second and each successive keystroke	0.50	0.85
E	Enter	Press the enter key	1.20	2.04
F	Function keys or shift	Press the function keys or shift		
M	Mental	Make a decision	1.50	2.55
S	Search	Search display for something (e.g., for cell in Excel sheet)	2.30	3.91
Rs	Response of system-scroll	to scroll one line (too small to matter, so assume 0)	0.00	0.00
Rm	Response of system-new menu	for new menu to appear; otherwise measure it	0.50	0.50

Note: The keystroke times shown in above include the time to move between keys.
Source: Developed from Society of Automotive Engineers (2002).

```
wait until list of streets to appear, then
1: read highlighted item, decide if 1st character of name of current entry matches desired street
2: if first pass, then select method (**sequential scroll or alpha scroll by 1st letter of name)
        if not matching 1st character then
                press right arrow key (**to go to next character)
                go to step 1 (**to read the next item)
        if matching 1st character then
                current character = second character
                (**compare current character of highlighted item with desired street)
                2.1 if (current character matches and current character is last character) then
                        confirm correct entry (**mental operation**)
                        hit enter
                        go to next subgoal
                if (current character matches and current character is not last character) then
                        increment current character (**add 1 to character index**)
                        go to 2.1 (**compare current character )
                if current character does not match displayed character then
                        press key to scroll down 1 entry
                        go to 2.1
exit: next subgoal
```

Figure 40.2 Pseudocode example.

Though many of these assumptions are not true, adjustments can be made for them, and often the adjustments are small. Furthermore, violations of assumptions tend to affect all interfaces equally, so decisions about which of several interfaces is best still hold. As a practical matter, the estimates are good enough for most engineering decisions. Readers should keep in mind that J2364 only requires the use of 10 subjects at most, so there is some error in those estimates. Those errors are likely to be as large as variability among analysts and among J2365 estimates.

Although SAE J2365 is useful, it needs to be updated to include time estimates for elements that occur for contemporary interfaces that were not included in the original version. They include drag, scroll, flick, search a list, turn a knob, push and hold, and write. Data to develop some of those elements exist (e.g. Green, Kang, & Lin, 2015; Schneegaß, Pfleging, Kern, & Schmidt, 2011).

Pettitt's occlusion time calculation

In his dissertation, Pettitt (2007) (see also Pettitt, Burnett, & Stevens, 2007, for one of several summaries) developed a related method for estimating occlusion task times. As was described earlier, in the occlusion procedure (the NHTSA guidelines—U.S. Department of Transportation, 2013); ISO Standard 16673—International Organization for Standardization, 2007; SAE Recommended Practice J2364—Society of Automotive Engineers, 2004a), subjects are allowed to look intermittently at a display while performing a task, simulating looking back and forth between the road scene and an in-vehicle system.

For all practical purposes, Pettitt's method is an extension of SAE J2365, with additional rules to determine what can be done during occlusion intervals.

- Rule 1: During the vision interval (assumed to be 1.5 s), the task progresses normally without interruption.
- Rule 2: Operators that do not require vision (e.g., mental) begun during vision can continue during occlusion.
- Rule 3: Only operators that do not require vision can begin during occlusion (e.g., a key can be pressed if the finger is already resting there).

Actually, Pettit uses fewer elemental times than in J2365 and except for reaching for the device (Rf = 0.31 s), the times are slightly different (M = 1.25 s, H = home / move to key = 0.62 s, K = press a key = 0.2 s).

Using these methods, Pettitt, Burnette, and Stevens reported estimates to be 10% greater (range 2–22%) for static task time (estimate for six tasks from 12 drivers) and 13% greater (range 2 to 12%) for total shutter open time in an occlusion procedure, a reasonable approximation for engineering estimates. Keep in mind that the data from drivers is not the true time, just another estimate.

Beyond simple computational models, more process-oriented models, such as CogTool (Teo & John, 2008) and Distract-R (Salvucci, Zuber, Beregovaia, & Markley, 2005) have been developed but their use to solve practical problems has been limited.

What Changes to the Driver Interface are Likely in the Near Term?

Predicting the future is difficult to make with any accuracy (Blank, 2015; Malone, 1997; Tetlock & Gardner, 2015), but several themes seem reasonable. To some degree, the pace at which advanced technology is introduced by the motor vehicle industry depends on vehicle sales, and sales depend on the state of the economy and the price of gasoline, which in turn depends on the price of crude oil (Swanson, 2016). However, there are a number of important themes that will influence which technologies will be introduced and their success.

Motor-vehicle automation

There was an effort in the 1990s to foster the development of automated vehicles (Congress, 1994)—an effort largely driven by the U.S. and other governments, not the private sector. The technology was not ready. Now with advances from the DARPA Grand Challenge (Chow, 2014; Thrun et al., 2006), demonstrations by Google and others, the efforts of Tesla, and statements by auto company leaders (Nissan, 2014), more automated and maybe fully automated vehicles will be produced in the near term. Some suggest full automation will occur in 2019 or 2020 (http://www.driverless-future.com/?page_id=384; Litman, 2014) and the auto suppliers and manufacturers are racing to meet those goals. Usually, those estimates refer to when a first vehicle will appear on the market, not when there will be widespread market penetration. Google and Tesla are pushing for that to occur sooner, but regulatory and insurance impediments could be just as important as technology development. See, for example, Ravid (2014).

Automation of the driving task has occurred in steps, first over speed (cruise control), then over lane choice, turning at intersections, and so forth. As the level of automation increases, the driver does less and the vehicle does more. Table 40.9

Table 40.9 Schemes to categorize levels of automation.

NHTSA Level	*Description*	*BASt Level*
0—none	Traditional, manually driven vehicles, including those with automated warnings or automated secondary controls (e.g., headlights, turn signals, vehicles before 1971)	driver only
1—function specific	At least one independent automated primary control function (e.g., steering, braking) such as adaptive cruise control and electronic stability control	assisted
2—combined function	Two or more automated primary functions that work in unison (e.g., adaptive cruise control with lane centering). The driver can retake control without warning	partially automated
3—limited self-driving	Driver can cede full control of the vehicle in some situations (garage parking pilot). The driver can retake control after a warning and transition period.	highly automated
4—full self-driving	Driver has no expectation to be able to resume control (e.g., robot taxi).	fully automated

shows some of the schemes used to categorize the levels of automation. In addition, the Society of Automotive Engineers has a classification scheme that maps onto these fairly well. It has six levels (0–5), with their scheme making some additional distinctions at the high levels of automation. See SAE Information Report J3016 (Society of Automotive Engineers, 2014). There are other taxonomies as well (Parasuraman, Sheridan, & Wickens, 2000). Some believe that parts of Step 3 may be skipped because of transfer of control risks (Oremus, 2016).

Another perspective, and a potentially more useful one, specifies exactly what the vehicle can do, what Nowakowski, Shladover, and Chan (2015) refer to a behavioral competency (Figure 40.3). Readers may also find Marinik et al. (2014) and Trimble, Bishop, Morgan, and Blanco (2014) to be helpful.

Automation is not only important for cars but for trucks as well, in particular for truck platoons, with a human driver in the lead vehicle and a human driver supported by automation or no human driver in the followers (Bishop, Bevly, Switkes, & Park, 2014; Robinson, Chan, & Coelingh, 2010).

Distraction and fatigue warnings and takeover

Until now, efforts to detect if drivers are distracted or fatigued in real time have largely been research projects. The algorithms developed use driving performance data (e.g., steering wheel angle, gap), and may use facial expression, eye gaze data, hand movements, and other inducia to determine if drivers are distracted or fatigued (Dinges, Mallis, Maislin, & Powell, 1998; Dong, Hu, Uchimura, & Murayama, 2011; Kircher & Ahlstrom, 2009; Liang, Reyes, & Lee, 2007; Wang, Zhou, & Ying, 2010). This occurring because some of the data needed are already on the CAN bus, and automation may provide the capability to take over if needed.

Shared vehicle use

Shared use is being motivated by (a) the increased cost of parking in urban areas, (b) increased congestion and the desire to reduce the number of vehicles on the road, and (c) smartphone technology that supports requesting vehicles on demand (e.g., Uber, Lyft, Sidecar). What is unknown is if those vehicles will be manually driven or automated. If the use of shared vehicles becomes large enough, there is the potential for vehicle sales to decline (Martin, Shaheen, & Lidicker, 2010). There is also the possibility that shared vehicle use may lead to demands for driver and passenger interfaces that are quite different from those in current vehicles (much more rapidly synchable displays, shared displays). One major unknown, for both drivers and passengers, is who, for how long, and under what conditions people can stare at visual displays without experiencing motion sickness (Green, 2016).

Integration of smartphones

Driver interfaces take several years to develop because the hardware and software needs to be automotive grade—they need to last 20 years, always work, and not cause injury in the event of a crash. Consumer electronics have a 6-month cycle, and their failure to function is not critical to the user. However, the attractive feature set and

Behavioral Competency ✔ Required competency for SAE L3+ system ★ Required behaviors differs for SAE L4+ driverless ? Safety case uncertain for requirement OPT Competency may be optionally included in SAE L3/4	Area of Operation				
	Freeway	Rural Highway	City Streets	Valet Parking	Low-Speed Shuttles
Detect Operating Envelope Based on ODD Engagement / Disengagement Criteria	✔	✔	✔	✔	✔
Detect System Faults & Failures	✔	✔	✔	✔	✔
Transition to Minimal Risk State	L4+	L4+	L4+	✔	✔
Detect & Respond to Speed Limit Changes and Speed Advisories	✔	✔	✔		✔
Perform High Speed Merge (e.g., Freeway)	★	★			
Perform Lane Change / Lower Speed Merge	?	?	✔		
Move Out of the Travel Lane and Park (e.g., on to the Shoulder for Minimal Risk)	?	?	✔		?
Detect & Respond to Encroaching Oncoming Vehicles	?	✔	✔	✔	✔
Detect Passing and No Passing Zones & Perform Passing Maneuvers		OPT	★		
Perform Car Following (Including Stop & Go)	✔	✔	✔	✔	✔
Detect & Respond to Stopped Vehicles	✔	✔	✔	✔	✔
Detect & Respond to Lane Changes	✔	✔	✔		?
Detect & Respond to Static Obstacles in the Roadway	✔	✔	✔	✔	✔
Detect Bikes, Peds, Animals, etc.	?	✔	✔	✔	✔
Respond to Bikes, Peds, Animals, etc.	?	★	✔	✔	✔
Detect Traffic Signals & Stop/Yield Signs	✔	✔	✔		OPT
Respond to Traffic Signals & Stop/Yield Signs	★	★	✔		OPT
Navigate Intersections & Perform Turns		?	✔	✔	✔
Navigate a Parking Lot & Locate Spaces			OPT	✔	
Detect & Respond to Access Restrictions (One-Way, No Turn, Ramps, etc.)	★	?	✔	✔	✔
Detect Work Zones and/or Safety Officials	✔	✔	✔		?
Navigate Work Zones and/or Safety Officials	★	★	★		?
Detect Emergency Vehicles	✔	✔	✔		✔
Respond to Emergency Vehicles	★	★	★		✔

Figure 40.3 Behavioral competency requirements. *Source:* Nowakowski, Shladover, Chan, and Tan (2015).

low cost of Apple CarPlay and Android Auto has led to the integration of smartphones into motor vehicles (Shelly, 2015). There is a huge reluctance on the auto manufacturers to allow Apple and Google into the car because of the considerable value of the data collected, data the manufacturers want to own. One possible future is that the manufacturer's physical interface disappears, with the user expected to provide a smartphone or tablet for that purpose.

Larger programmable displays

Counter to the no-physical-interface future is the prospect of larger auto manufacturer-provided displays. Over time, the size of these displays has increased both in the center stack and the speedometer / tachometer cluster. Currently, the largest display is the 17-inch diagonal LCD in the Tesla models S and X, though other companies, such as Volvo are moving in that direction. Growth in the size of the speedometer / tachometer display should continue, with the display getting wider.

Larger head-up displays (HUDs) and augmented reality

Head-up displays, to a large degree, have been a technology in search of an application. To date, the most common use for them has been as a speedometer. However, there are few instances where the driver needs to instantly glance at the speedometer and the fraction of a second that a HUD saves is often not that critical. Most glances to the speedometer are discretionary. However, the driver does need to make navigation-related split second decisions regarding lane choice and turns at intersections, especially when the workload is high. In those instances, a HUD close to the line of sight is very beneficial. Lately, research has begun to explore the advantages of full-windshield HUDs for collision avoidance (Lin, Kang, & Green, 2016), another split-second decision.

Connectivity

There are significant potential safety benefits of cars being wirelessly connected to each other via dedicated short range communication (DSRC) and sending SAE J2735 messages to each other about their location, heading, speed, and other information (Kenney, 2011). The information would be otherwise provided by cameras, radar, LIDAR, and other sensors at a potentially greater cost. However, to provide the desired information for automated vehicles, these on-board sensors may be needed anyway.

Where can one Learn More on Driver Interface Design and Evaluation?

For those seeking additional information, there are three, high quality, scientific conferences focusing on driver interface research is presented (Table 40.10). These three conferences typically last about three days, have about 200–250 attendees, and focus not just on high-quality research presentations but on providing opportunities for attendees to interact. The chapter author has attended the first two conferences and is the chair for the first one (all listed in alphabetical order). Similar material is presented at the Human Factors and Ergonomics Society Annual meeting in the surface

Table 40.10 Focused conferences on motor vehicle driver interfaces.

Conference	*When*	*Comments*
Automotive User Interfaces and Interactive Vehicular Applications Conference (Auto-UI.org)	Alternates between the United States and Europe; held in fall	Day 1—free workshops that are attended by almost all attendees
Driving Assessment conference (http://drivingassessment.uiowa.edu)	Biannual, midsummer, often at a ski resort so the family can vacation at the same time, always in the United States	All aspects of driving, not just driver interface
Driver Distraction and Inattention (https://ddi2017.sciencesconf.org/; http://www.chalmers.se/hosted/ddi2013-en/)	Held in Sweden and Australia	

transportation sessions (HFES.org). That meeting is larger, longer, and covers all aspects of human factors, not just driving. There are also useful sessions at the Society of Automotive Engineers Annual Congress (SAE.org).

In addition, there are several commercial conferences sponsored by the International Quality and Productivity Center (IQPC.com—Automotive Cockpit HMI), we-CONECT (we-conect.com Car HMI Concepts and Systems), and TU-Automotive (tu-auto.com, various names—search for HMI), whose location varies. These conferences involve only invited speakers and tend not to have published proceedings. Their fees are often double those of the scientific conferences. The author has chaired conference sponsored by IQPC and we-CONECT in the Detroit area. The first two conferences are small (about 100 attendees) and give a high priority to meeting people. The TU-automotive meeting in Detroit is reported to attract more than 3000 people, but not many researchers. Its focus is on automotive technology in general, not just the human interface. The focus at these commercial meetings is more on the results of research than the research per se.

The author's personal recommendation is that anyone engaged in driver interface development or research should go to at least one of the scientific conferences each year. If they are new to the field, then the commercial conferences will be beneficial.

Closing Thoughts

This chapter makes the following key points:

- Driving is quite different from sitting at a desk in an office because of the concern for crash risk. People die while driving—lots of them.
- Nonetheless, driver interface design should follow the same golden Gould and Lewis principles used to design ordinary office applications—(a) early focus on users and tasks, (b) empirical measurement, and (c) iterative design.

- The best data on crashes comes from three U.S. databases that are online and available for free to anyone: (a) the Fatality Analysis Reporting System (FARS), (b) the National Automotive Sampling System (NASS) General Estimates System (GES), and (c) the Crashworthiness Data System (CDS). There are no foreign data bases that are equally accessible or as comprehensive.
- When designing interfaces, one must consider how drivers are licensed (which varies widely between countries), the driving culture, the compliance with traffic regulations (also highly variable) and the driver demographics. In the United States, even elderly drivers still drive. Demographic data on drivers outside the United States is limited, with data on Chinese drivers being a special need.
- Not just cars, but all types of motor vehicles need to be designed to accommodate drivers, passengers, and cargo (scooters, SUVs, minivans, light trucks, heavy trucks, buses). The largest market for motor vehicles is now China. In the United States, which has been the largest market for a long time, and equal number of cars and trucks have been sold, due in part to the popularity of pickup trucks and SUVs. Worldwide, commercial vehicles continue to make up a large fraction of the world's fleet.
- Trips are made for a wide variety of purposes, which need to be considered in assessing driver interfaces.
- In-vehicle systems and interfaces that have received consideration attention include smartphones, navigation systems, menu interfaces, speech interfaces, and workload managers.
- When assessing user performance with in-vehicle tasks, one also collects primary task measures including measures of longitudinal control (e.g., standard deviation of lateral position, the number of lane departures), lateral control (e.g., gap, mean time to collision), and glances (e.g., percentage of off-road glances greater than 2 s), as well as other measures. It is essential that measures be identified as per SAE Recommended Practice J2944. It is common for there to be 30–50 dependent measures in a driving simulator study, and hundreds in an on-the-road study. A major challenge in motor vehicle interface evaluation is dealing with performance tradeoffs between measures.
- Over the last few years, there have been significant advances in driving simulators and instrumented vehicles, which have improved their quality and reduced their cost for safety and usability evaluations. There are some low-cost options (e.g., OpenDS).
- The key design and evaluation documents are the NHTSA guidelines for distraction, a long list of ISO documents for all aspects of safety and usability, similar SAE documents and to a lesser extent, other guidelines. Of the test procedures identified, the NHTSA occlusion procedure is the most important. For estimating performance in those procedures SAE J2365 is most important along with Pettit's method to predict occlusion time. Enhancements to SAE J2365 to include touch screen elements, mostly gestures, are needed.
- Likely changes in driving interfaces in the future are most likely to be influenced by motor vehicle automation, shared vehicle use, integration of smartphones, larger programmable displays, larger HUDS and augmented reality, and connectivity.
- For those interested in the topic of driver interfaces, this chapter is just a beginning. They should read the scientific literature, participate in professional societies, and attend their conferences. Key scientific conferences, sources of new information,

include the Automotive User Interface (AutoUI), the Human Factors Annual Conference, the Driver Distraction Conference, and the SAE Annual Congress, listed in order of importance. Commercial conferences such as those run by IQPC and we-CONECT are also of value.

Thus, although the HCI literature provides a framework for test methods and evaluation, a great deal is specific to motor-vehicle interfaces because of the safety-critical nature of the context and the timesharing not found in office activities. To meet the needs of the future, the cost of the methods needs to be reduced, and reliable tools, especially for recording gaze, are needed. Significant research is needed to support the development of driver performance models (and workload managers) and understand how drivers use real driver interfaces.

References

Akamatsu, M., Green, P. A., & Bengler, K. (2013). Automotive technology and human factors research: Past, present, and future. *International Journal of Vehicular Technology*, Article ID 526180.

Alliance of Automobile Manufacturers. (2006). *Statement of principles on human-machine interfaces (HMI) for in-vehicle information and communication systems* (version 3). Washington, DC: Author.

Anderson, M. (2015, October 29). Technology device ownership: 2015. *Pew Research Center*. Retrieved from http://www.pewinternet.org/2015/10/29/technology-device-ownership-2015/

Automakers trail in the great gizmo race. (2011, June 10). *Road Show*. Retrieved from http://www.cnet.com/roadshow/news/automakers-trail-in-the-great-gizmo-race/

Average life of a desktop PC 2012-2017. (2016). *Statista*. Retrieved from http://www.statista.com/statistics/267465/average-desktop-pc-lifespan/

Baltodano, S., Sibi, S., Martelaro, N., Gowda, N., & Ju, W. (2015, September). The RRADS platform: A real road autonomous driving simulator. In *Proceedings of the Seventh International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 281–288). New York, NY: ACM.

Bao, S., Flannagan, C., Xiong, H., & Sayer, J. (2014). Eye glance behavior associated with cell-phone use examination with naturalistic driving data. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 2112–2116). Iowa City, IA: University of Iowa.

Barón, A., & Green, P. (2006). *Safety and usability of speech interfaces for in-vehicle tasks while driving: A brief literature review* (Report No. UMTRI-2006-5). Ann Arbor, MI: University of Michigan Transportation Research Institute.

Bayly, M., Young, K. L., & Regan, M. A. (2008). Sources of distraction inside the vehicle and their effects on driving performance. In M. A. Regan, J. D. Lee, & K. Young (Eds.), *Driver distraction: Theory, effects, and mitigation* (pp. 192–210). Boca Raton, FL: CRC Press.

Bertrand, M., Djankov, S., Hanna, R., & Mullainathan, S. (2008). Corruption in driving licensing process in Delhi. *Economic and Political Weekly, 43*(5), 71–76.

Bhise, V. D. (2011). *Ergonomics in the automotive design process.* Boca Raton, FL: CRC Press.

Bishop, R., Bevly, D., Switkes, J., & Park, L. (2014). Results of initial test and evaluation of a driver-assistive truck platooning prototype. In *2014 IEEE Intelligent Vehicles Symposium Proceedings* (pp. 208–213). New York, NY: IEEE.

Blank, S. (2015). Blank's Rule—To predict the future 1/3 of you need to be crazy. Retrieved from http://steveblank.com/2015/12/15/blanks-rule-to-predict-the-future-13-of-you-need-to-be-crazy/

Boyle, L. N., Lee, J. D., Peng, Y., Ghazizadeh, M., Wu, Y., Miller, E., & Jenness, J. (2013, August). *Text reading and text input assessment in support of the NHTSA visual-manual driver distraction guidelines* (Report No. DOT HS 811 820). Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.

Bureau of Transportation Statistics. (2014). *Number of U.S. aircraft, vehicles, vessels, and other conveyances.* Retrieved from http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_statistics/html/table_01_11.html

Caird, J. K., Scialfa, C. T., Ho, G., & Smiley, A. (2006). A meta-analysis of driving performance and crash risk associated with the use of cellular telephones while driving (pp. 478–485). In *Proceedings of the Third International Driving Symposium on Human factors in Driver Assessment, Training and Vehicle Design*, Iowa City, IA: University of Iowa.

Campbell, J., Carney, C., & Kantowitz, B. H. (1997). *Human factors design guidelines for advanced traveler information systems (ATIS) and commercial vehicle operations (CVO)* (Report No. FHWA-RD-98-057). Washington, DC: U.S. Department of Transportation, Federal Highway Administration.

Campbell, J., Doerzaph, Z. R., Richard, C. M., & Bacon, L.P. (2014). Human factors design principles for the driver-vehicle interface (DVI). In *Sixth International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 1–6). New York, NY: ACM.

Campbell, K. L. (2012). The SHRP 2 naturalistic driving study: Addressing driver performance and behavior in traffic safety. *TR News*, *282*, 30–35.

Card, S. K., Moran, T. P., & Newell, A. (1980, July 23). The keystroke-level model for user performance time with interactive systems. *Communications of the ACM*, 7, 396–410.

Card, S. K., Moran, T. P., & Newell, A. (1983). *The psychology of human-computer interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cars Move America. (2016). Retrieved from https://autoalliance.org/resources/cars-move-america/

CBC (2010, January 15). GPS distraction. Retrieved from http://www.cbc.ca/marketplace/episodes/2010-episodes/gps-distraction

Chang, J. C., Lien, A., Lathrop, B., & Hees, H. (2009). Usability evaluation of a Volkswagen group in-vehicle speech system. In *Proceedings of the First International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI2009)* (pp. 137–144). Essen, Germany. New York: ACM.

Chen, B. X. (2016, January 27). Siri, Alexa and other virtual assistants put to the test. *New York Times*. Retrieved from http://www.nytimes.com/2016/01/28/technology/personaltech/siri-alexa-and-other-virtual-assistants-put-to-the-test.html?_r=1

Chisholm, S. L., Caird, J. K., & Lockhart, J. (2008). The effects of practice with MP3 players on driving performance, *Accident Analysis and Prevention*, *40*(2), 704–713.

Chow, D. (2014, March 21). DARPA and drone cars: How the U.S. military spawned self-driving car revolution. *Live Science*. Retrieved from http://www.livescience.com/44272-darpa-self-driving-car-revolution.html

Collet, C., Guillot, A., & Petit, C. (2010a). Phoning while driving I: A review of epidemiological, psychological, behavioural and physiological studies. *Ergonomics*, *53*(5), 589–601.

Collet, C., Guillot, A., & Petit, C. (2010b). Phoning while driving II: a review of driving conditions influence. *Ergonomics*, *53*(5), 602– 616.

Commission of the European Communities. (2008). Commission recommendation of 26 May 2008 on the safe and efficient in-vehicle information and communication systems: Update of the European Statement of Principles on human-machine interface. *EUR-Lex*. Retrieved from http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32008H0653

Congress, N. (1994). The automated highway system: An idea whose time has come. *Public Roads*, *58*(1), http://www.fhwa.dot.gov/publications/publicroads/94summer/p94su1.cfm

Davis, S. C., Diegel, S. W., & Boundy, R. G. (2015). *Transportation energy data book* (34 ed.). Oak Ridge, TN: U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, Vehicle Technologies Office.

Dinges, D. F., Mallis, M. M., Maislin, G., & Powell, I. V. (1998). *Evaluation of techniques for ocular measurement as an index of fatigue and the basis for alertness management* (report DOT HS-808 762), Washington, D.C.: U.S. Department of Transportation.

Dingus, T. A., Guo, F., Lee, S., Antin, J. F., Perez, M., Buchanan-King, M., & Hankey, J. (2016). Driver crash risk factors and prevalence evaluation using naturalistic driving data. In *Proceedings of the National Academy of Sciences 2015* (pp. 2636–2641). Washington, DC: National Academy of Sciences.

Dingus, T. A., Hanowski, R. J., & Klauer, S. G. (2011). Estimating crash risk. *Ergonomics in Design*, *19*(4), 8–12.

Dingus, T. A. & Klauer, S. G. (2008). *The relative risks of secondary task induced driver distraction* (SAE paper 2008-21-0001). Warrendale, PA: Society of Automotive Engineers.

Dingus, T. A., Klauer, S. G., Neale, V. L., Petersen, A., Lee, S. E., Sudweeks, J. D.,…Bucher, C. (2006). *The 100-car naturalistic driving study, Phase II-results of the 100-car field experiment (Report No. HS-810 593)*. Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.

Dong, Y., Hu, Z., Uchimura, K., & Murayama, N. (2011). Driver inattention monitoring system for intelligent vehicles: A review. *IEEE Transactions on Intelligent Transportation Systems*, *12*(2), 596–614.

Dopart, C., Häggman, A., Thornberry, C., Mehler, B., Dobres, J., & Reimer, B. (2013, September). A driving simulation study examining destination entry with iOS 5 Google Maps and a Garmin Portable GPS System. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (pp. 1889–1893). Santa Monica, CA: Human Factors and Ergonomics Society.

Driven to distraction. (2010, July 19). *New York Times*. Retrieved from http://topics.nytimes.com/top/news/technology/series/driven_to_distraction/index.html

Ervin, R., Sayer, J., LeBlanc, D., Bogard, S., Mefford, M., Hagan, M.,…Winkler, C. (2005). *Automotive collision avoidance system field operational test report: methodology and results.* (Report No. HS-809 900). Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.

Farber, E., Blanco, M., Foley, J. P., Curry, R., Greenberg, J., & Serafin, C. (2000). Surrogate measures of visual demand while driving. In *Proceedings of the IEA/HFES 2000 Congress* [CD-ROM]. Santa Monica, CA: Human Factors and Ergonomics Society.

Farmer, C. M., Braitman, K. A., & Lund, A. K. (2010). *Cell phone use while driving and attributable crash risk*. Arlington, VA: Insurance Institute for Highway Safety.

Flannagan, C. A., Bao, S., & Klinich, K. D. (2012). *Driver distraction from cell phone use and potential for self-limiting behavior* (Report No. UMTRI-2012-36). Ann Arbor, MI: University of Michigan Transportation Research Institute.

Garay-Vega, L., Pradhan, A. K., Weinberg, G., Schmidt-Nielsen, B., Harsham, B., Shen, Y.,…Fisher, D. L. (2010). Evaluation of different speech and touch interfaces to in-vehicle music retrieval systems. *Accident Analysis and Prevention*, *42*(3), 913–920.

General Motors Corporation. (2005). *Automotive Collision Avoidance System Field Operational Test (ACAS FOT) final program report* (Report No. DOT HS 809 886). Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.

Go, E., Morton, A., Famewo, J., & Angel, H. (2006). *Final report: Evaluation of industry safety principles for in-vehicle information and communication systems*, Ottawa, Canada: Transport Canada.

Gould, J. D., & Lewis, C. (1985). Designing for usability: Key principles and what designers think. *Communications of the ACM*, *28*(3), 300–311.

Green, P. (1995a). Automotive techniques. In J. Weimer (Ed.), *Research techniques in human engineering* (2nd ed., pp. 165–208). New York, NY: Prentice-Hall.

Green, P. (1995b). *Measures and methods used to assess the safety and usability of driver information systems* (Report No. FHWA-RD-94-088). McLean, VA: U.S. Department of Transportation, Federal Highway Administration.

Green, P. (1995c). *Suggested procedures and acceptance limits for assessing the safety and ease of use of driver information systems* (Report No. FHWA-RD-94-089). McLean, VA: U.S. Department of Transportation, Federal Highway Administration.

Green, P. (1999a). Estimating compliance with the 15-second rule for driver-interface usability and safety. In *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting* (pp. 165–208). Santa Monica, CA: Human Factors and Ergonomics Society.

Green, P. (1999b). *Navigation system data entry: Estimation of task times* (Report No. UMTRI-99-17). Ann Arbor, MI: University of Michigan, Transportation Research Institute.

Green, P. (1999c). The 15-second rule for driver information systems. In *ITS America Ninth Annual Meeting Conference Proceedings* (pp. 1–9). Washington, DC: Intelligent Transportation Society of America.

Green, P. (2000a, June 8). *The human interface for ITS display and control systems: developing international standards to promote safety and usability.* Paper presented at the International Workshop on ITS Human Interface, Utsu, Japan.

Green, P. (2000b). Crashes induced by driver information systems and what can be done to reduce them (SAE paper 2000-01-C008). In *Convergence 2000 Conference Proceedings* (pp. 27–36), Warrendale, PA: Society of Automotive Engineers.

Green, P. (2001a, February 5). *Safeguards for on-board wireless communications.* Paper presented at the Second Annual Plastics in Automotive Safety Conference, Troy, Michigan.

Green, P. (2001b). *Synopsis of driver interface standards and guidelines for telematics as of mid-2001* (Report No. UMTRI-2001-23). Ann Arbor, MI: University of Michigan, Transportation Research Institute.

Green, P. (2001c, February 19–20). *Variations in task performance between younger and older drivers: UMTRI research on telematics.* Paper presented at the Association for the Advancement of Automotive Medicine Conference on Aging and Driving, Southfield, MI.

Green, P. (2004). Driver distraction, telematics design, and workload managers: Safety issues and solutions (SAE paper 2004-21-0022). In *Proceedings of the 2004 International Congress on Transportation Electronics* (pp. 165–180). Warrendale, PA: Society of Automotive Engineers.

Green, P. (2012). Using standards to improve the replicability and applicability of driver interface research. In *Proceedings of the Fourth International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 15–22). New York, NY: ACM.

Green, P. (2013). Standard definitions for driving measures and statistics: overview and status of recommended practice J2944. In *Proceedings of the Fifth International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 184–191). New York, NY: ACM.

Green, P. (2016). *Motion sickness and concerns for self-driving vehicles* (Report No. UMTRI-2016-*). Ann Arbor, MI: University of Michigan Transportation Research Institute.

Green, P., George, J., & Jacob, R. (2003). *What constitutes a typical cell phone call* (Report No. UMTRI-2003-38). Ann Arbor, MI: University of Michigan Transportation Research Institute.

Green, P., Kang, T.-P., & Lin, B. T. W. (2015). Touch-screen task-element times for improving SAE Recommended Practice J2365: A first proposal (Report No. ATLAS-2015-07). Ann Arbor, MI: University of Michigan Transportation Research Institute.

Green, P., Levison, W., Paelke, G., & Serafin, C. (1993). Preliminary human factors guidelines for driver information systems (Report No. UMTRI-93-21). Ann Arbor, MI: The University of Michigan Transportation Research Institute.

Green, P., & Park, J.-S. (2013). Evaluation of a navigation radio using the think-aloud method. *International Journal of Vehicular Technology*, *2013*, Article ID 705086.

Highway Loss Data Institute. (2009). Hand-held cell phone laws and collision claim frequencies, *Highway Data Loss Institute Bulletin*, *26*(17), 1–5.

Highway Loss Data Institute. (n.d. a). *Older drivers*. Retrieved April 3, 2007, from http://www.iihs.org/iihs/topics/laws/olderdrivers

Highway Loss Data Institute. (n.d. b). *Teenagers*. Retrieved from http://www.iihs.org/iihs/topics/laws/graduatedlicensestatelaws?topicName=teenagers

Hoedemaeker, M., de Ridder, S. N., & Janssen, W. H. (2002). *Review of European human factors research on adaptive interface technologies for automobiles* (TNO report TM-02-C031). Soesterberg, Netherlands: TNO Institute for Perception.

Hornbæk, K., & Law, E. L. C. (2007, April). Meta-analysis of correlations among usability measures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 617–626). New York: ACM.

Horrey, W. J., & Wickens, C. D. (2006). Examining the impact of cell phone conversations on driving using meta-analytic techniques. *Human Factors*, *48*(1), 196–205.

Hu, P. S., & Reuscher, T. R. (2004). *Summary of travel trends: 2001 national household travel survey* (technical report). Washington, DC: US. Department of Transportation, Federal Highway Administration.

International Organization for Standardization. (2007). *List of technical committees: TC 22/SC 13 Ergonomics applicable to road vehicles*. Geneva, Switzerland.

Jamson, S., Wardman, M., Batley, R., & Carsten, O. (2008). Developing a driving safety index using a Delphi state preference experiment. *Accident Analysis and Prevention*, *40*, 435–442.

Japan Automobile Manufacturers Association. (2004). *JAMA guideline for in-vehicle display systems, version 3.0*. Tokyo, Japan: Author.

Jeong, H., & Green, P. (2013). *SAE and ISO Standards for warnings and other driver interfaces elements: A summary* (Report No. UMTRI-2013-16). Ann Arbor, MI: University of Michigan Transportation Research Institute.

Johansson, E., Engstrom, J., Cherri, C., Nodari, E., Toffetti, A., Schindhelm, R., & Gelau, C. (2004). *Review of existing techniques and metrics for IVIS and ADAS assessment* (AIDE deliverable 2.2.1). Retrieved from http://www.aide-eu.org/pdf/sp2_deliv_new/aide_d2_2_1.pdf

Johnson, D. A., & Trivedi, M. M. (2011). Driving style recognition using a smartphone as a sensor platform. *2011 14th International IEEE Conference on Intelligent Transportation Systems (ITSC)* (pp. 1609–1615). New York, NY: IEEE.

Kelly, H. (2015, July 28). Which is the best digital assistant: Siri, Cortana, Alexa or Google Now? *CNN Money*. Retrieved from http://money.cnn.com/2015/07/28/technology/digital-assistant-interview/

Kenney, J. B. (2011). Dedicated short-range communications (DSRC) standards in the United States. *Proceedings of the IEEE*, *99*(7), 1162–1182.

Kidd, D. G., Dobres, J., Reagan, I. J., Mehler, B., & Reimer, B. (2015). *Considering visual-manual tasks performed during highway driving in the context of two different sets of guidelines for embedded in-vehicle electronic systems* (technical report). Arlington, VA: Insurance Institute for Highway Safety.

Kircher, K., & Ahlstrom, C. (2009, September). Issues related to the driver distraction detection algorithm AttenD. Paper presented at the First International Conference on Driver Distraction and Inattention (DDI 2009), Gothenburg, Sweden.

Klauer, S. G., Guo, F., Simons-Morton, B. G., Ouimet, M. C., Lee, S. E., & Dingus, T. A. (2014). Distracted driving and risk of road crashes among novice and experienced drivers. *New England Journal of Medicine*, *370*(1), 54–59.

Kolash, K. (2014). *Injury facts statistical highlights.* Chicago, IL: National Safety Council.

Koushki, P. A., Ali, S. Y., & Al-Saleh, O. I. (1999). Driving and using mobile phones: Impacts on road accidents. *Transportation Research Record*, *1694*, 27–33.

Kujala, T., Lesch, A., & Makela, J. (2014). Critical analysis on the NHTSA acceptance criteria for in-vehicle electronic devices. In *Sixth International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 1–8). New York, NY: ACM.

Lai, F. C. H. (2005). *Driver attentional demand to dual task performance* (Unpublished doctoral dissertation). University of Leeds, Institute for Transport Studies, Leeds.

Lai, J., Cheng, K., Green, P., & Tsimhoni, O. (2001, March). On the road and on the Web? Comprehension of synthetic and human speech while driving. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems* (pp. 206–212). New York, NY: ACM.

LeBlanc, D., Goodsell, R., Bareket, Z., Hagan, M. R. Buonarosa, M. L., Devonshire, J.,… Sayer, J. R. (2006). *Road departure crash warning system field operational test: Methodology and results. Volume 1: Technical report* (Report No. UMTRI-2006-9-1). Ann Arbor, MI: University of Michigan Transportation Research Institute.

Lee, J., Forlizzi, J., & Hudson, S. E. (2008). Iterative design of MOVE: A situationally appropriate vehicle navigation system. *International Journal of Human-Computer Studies*, *66*, 198–215.

Liang, Y., Lee, J. D., & Yekhshatyan, L. (2012). How dangerous is looking away from the road? Algorithms predict crash risk from glance patterns in naturalistic driving. *Human Factors*, *54*(6), 1104–1116.

Liang, Y., Reyes, M. L., & Lee, J. D. (2007). Real-time detection of driver cognitive distraction using support vector machines. *IEEE Transactions on Intelligent Transportation Systems*, *8*(2), 340–350.

Lin, B. T. W., Kang, T-P., & Green, P. (2016). *Drivers' responses to augmented reality warnings for crash scenarios at urban signalized intersections* (Report No. 2016-*). Ann Arbor, MI: University of Michigan Transportation Research Institute.

Litman, T. (2014). *Autonomous vehicle implementation predictions* (Report). Victoria, BC, Canada: Victoria Transport Policy Institute. Retrieved from http://www.vtpi.org/avip.pdf

Lo, E. (2013). *A model to predict driver task performance when interacting with in-vehicle speech interfaces for destination entry and music selection* (Doctoral dissertation). Ann Arbor, MI: University of Michigan.

Lo, V., & Green, P. A. (2013). Development and evaluation of automotive speech interfaces: Useful information from the human factors and the related literature. *International Journal of Vehicular Technology*, *2013*, Article ID 924170.

Lo, V., Green, P. A., & Franzblau, A. (2011). Where do people drive? Navigation system use by typical drivers and auto experts. *Journal of Navigation*, *64*(02), 357–373.

Low Vision Resources Center. (2003, June 6). *State vision screening and standards for license to drive.* Retrieved from http://lowvision.preventblindness.org/daily-living-2/state-vision-screening-and-standards-for-license-to-drive/

Maciej, J., & Vollrath, M. (2009). Comparison of manual vs. speech-based interaction with in-vehicle information systems. *Accident Analysis and Prevention*, *41*(5), 924–930.

Malone, J. (1997). *Predicting the future: From Verne to Bill Gates.* New York, NY: M. Evans.

Manes, D., Green, P., & Hunter, D. (1998). *Prediction of destination entry and retrieval times using keystroke-level models* (Report No. UMTRI–96–37). Ann Arbor, MI: University of Michigan, Transportation Research Institute.

Marinik, A., Bishop, R., Fitchett, V., Morgan, J. F., Trimble, T. E., & Blanco, M. (2014). *Human factors evaluation of level 2 and level 3 automated driving concepts: Concepts of operation* (Report No. DOT HS 812 044). Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.

Martin, E., Shaheen, S., & Lidicker, J. (2010). Impact of carsharing on household vehicle holdings: Results from North American shared-use vehicle survey. *Transportation Research Record*, *2143*, 150–158.

Mayhew, D. R., Simpson, H. M., & Pak, A. (2003). Changes in collision rates among novice drivers during the first months of driving. *Accident Analysis and Prevention*, *35*(5), 683–691.

McCartt, A. T., Hellinga, L. A., & Bratiman, K. A. (2006). Cell phones and driving: Review of research. *Traffic Injury Prevention*, *7*, 89–106.

McEvoy, S. P, Stevenson, M. R., McCartt, A. T., Woodward, M., Haworth, C., Palamara, P., & Cercarelli, R. (2005, July 12). Role of mobile phones in motor vehicle crashes resulting in hospital attendance: A case-crossover study. *British Medical Journal*, *331*: 428. Retrieved from http://bmj.bmjjournals.com/cgi/content/abstract/bmj.38537.397512.55v1

Mehler, B., Kidd, D., Reimer, B., Reagan, I., Dobres, J., & McCartt, A. (2015). Multi-modal assessment of on-road demand of voice and manual phone calling and voice navigation entry across two embedded vehicle systems. *Ergonomics*, *59*(3), 1–24.

Michon, J. A. (Ed.). (1993). *Generic intelligent driver support*. London: Taylor & Francis.

Nass, C., Jonsson, I. M., Harris, H., Reaves, B., Endo, J., Brave, S., & Takayama, L. (2005). Improving automotive safety by pairing driver emotion and car voice emotion. *CHI'05 Extended Abstracts on Human Factors in Computing Systems* (pp. 1973–1976). New York, NY: ACM.

Nissan. (2014, July 16). CEO Carlos Ghosn's speech to the Foreign Correspondents' Club of Japan. Retrieved from http://nissannews.com/en-US/nissan/usa/releases/ceo-carlos-ghosn-s-speech-to-the-foreign-correspondents-club-of-japan

Nowakowski, C. Friedman, D., & Green, P. (2001). *Cell phone ring suppression and HUD caller ID: Effectiveness in reducing momentary driver distraction under varying workload levels* (Report No. 2001-29). Ann Arbor, MI: University of Michigan, Transportation Research Institute.

Nowakowski, C., Shladover, S. E., Chan, C. Y., & Tan, H. S. (2015). Development of California regulations to govern testing and operation of automated driving systems. *Transportation Research Record*, *2489*, 137–144.

Nowakowski, C., Utsui, Y., & Green, P. (2000). *Navigation system evaluation: the effects of driver workload and input devices on destination entry time and driving performance and their implications to the SAE recommended practice* (Report No. UMTRI-2000-20). Ann Arbor, MI, University of Michigan: Transportation Research Institute.

Olson, J. R., & Nilsen, E. (1987–1988). Analysis of the cognition involved in spreadsheet software interaction. *Human-Computer Interaction*, *3*, 309–349.

Oremus, W. (2016, July 6). Is autopilot a bad idea? Why Ford, Google, and Volvo and others think Tesla is wrong about automation. *Future Tense*. Retrieved from http://www.slate.com/articles/technology/future_tense/2016/07/is_tesla_s_style_of_autopilot_a_bad_idea_volvo_google_and_others_think_so.html

Owens, J. M., McLaughlin, S. B., & Sudweeks, J. (2010). *On-road comparison of driving performance measures when using handheld and voice-control interfaces for mobile phones and portable music players (SAE paper 2010-01-1036)*. Warrendale, PA: Society of Automotive Engineers.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, *30*(3), 286–297.

Pauzie, A. (2008). A method to assess the driver mental workload: The driving activity load index, *IET Intelligent Transport Systems*, *2*(4), 315–322.

Pettitt, M. (2007). *Development of an extended keystroke level model for predicting the visual demand of in-vehicle interfaces* (Unpublished PhD dissertation). Nottingham: University of Nottingham.

Pettitt, M., Burnett, G., & Stevens, A. (2007). An extended keystroke level model (KLM) for predicting the visual demand of in-vehicle information systems. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1515–1524). New York, NY: ACM.

Piechulla, W., Mayser, C., Gehrke, H., & König, W. (2003). Reducing drivers' mental workload by means of an adaptive man-machine interface. *Transportation Research Part F: Traffic Psychology and Behaviour*, *6*, 233–248.

Pournami, S., Large, D. R., Burnett, G., & Harvey, C. (2014). Comparing the NHTSA and ISO Occlusion Test protocols: How many participants are sufficient? In *Proceedings of the Seventh International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 110–116). New York, NY: Association for Computing Machinery.

Ranney, T. A. (2008). *Driver distraction: A review of the current state-of-knowledge* (Report No. DOT HS 810 787). Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.

Ranney, T. A., Baldwin, G. H. S., Parmer, E., Domeyer, J., Martin, J., & Mazzae, E. N. (2011). *Developing a test to measure distraction potential of in-vehicle information system tasks in production vehicles (Report No. DOT HS 811 463)*. Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.

Ravid, O. (2014). Don't sue me, I was just lawfully texting and drunk when my autonomous car crashed into you. *Southwest Law Review*, *44*, 175.

Redelmeier, D. A., & Tibshirani, R. J. (1997). Association between cellular-telephone calls and motor vehicle collisions. *New England Journal of Medicine*, *336*, 453–458.

Redelmeier, D. A., & Tibshirani, R. J. (2001). Car phones and car crashes: Some popular misconceptions. *Canadian Medical Association Journal*, *164*(11), 1581–1582.

Reimer, B., Mehler, B., Dobres, J., & Coughlin, J. F. (2013). The effects of a production level "voice-command" interface on driver behavior: Summary findings on reported workload, physiology, visual attention, and driving performance (Report No. 2013-17A). Cambridge, MA: MIT AgeLab.

Robinson, T., Chan, E., & Coelingh, E. (2010, October). *Operating platoons on public motorways: An introduction to the SARTRE platooning programme*. Paper presented at the Seventeenth World Congress on Intelligent Transport Systems, Busan, Korea.

Roskam, A. J., Brookhuis, K. A., deWaard, D., Carsten, O. M. J., Read, L., Jamson, S.,… Victor, T. (2002). *Development of experimental protocol* (HASTE Deliverable 1). Brussels, Belgium: European Commission.

Ross, T., Midtland, K., Fuchs, M., Pauzie, A., Engert, A., Duncan, B.,…May, A. (1996). *HARDIE design guidelines handbook: Human factors guidelines for information presentation by ATT Systems*. Luxembourg: Commission of the European Communities.

Salvucci, D. D., Zuber, M., Beregovaia, E., & Markley, D. (2005). Distract-R: Rapid prototyping and evaluation of in-vehicle interfaces. In *Proceedings of the SIGCHI Conference on Human factors in Computing Systems* (pp. 581–589). New York, NY.

Santos, A., McGuckin, N., Nakamoto, H. Y., Gray, D., & Liss, S. (2011). *Summary of travel trends: 2009 national household travel survey* (Report No. FHWA-PL-ll-022). Washington, DC: U.S. Department of Transportation, Federal Highway Administration.

Sayer, J., Devonshire, J., & Flannagan, C. (2005). *The effects of secondary tasks on naturalistic driving performance* (Report No. UMTRI-2005-29). Ann Arbor, Michigan: University of Michigan Transportation Research Institute.

Schindhelm, R., Gelau, C., Keinath, A., Bengler, K., Kussmann, H., Kompfner, P.,…Martinetto, M. (2004). *Report on the review of available guidelines and standards* (AIDE deliverable 4.3.1). Brussels, Belgium: European Commission. Retrieved from http://www.aide-eu.org/pdf/sp4_deliv_new/aide_d4_3_1.pdf

Schneegaß, S., Pfleging, B., Kern, D., & Schmidt, A. (2011, November). Support for modeling interaction with automotive user interfaces. In *Proceedings of the Third International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 71–78). New York, NY: ACM.

Schweitzer, J., & Green, P. A. (2007). *Task acceptability and workload of driving urban roads, highways, and Expressways: Ratings from video clips* (Report No. UMTRI-2006-6). Ann Arbor, MI: University of Michigan Transportation Research Institute.

Shelly, P. (2015). Addressing challenges in automotive connectivity: mobile devices, technologies, and the connected car. *SAE International Journal of Passenger Cars-Electronic and Electrical Systems*, *8*(2015-01-0224), 161–169.

Short, J. (2014). *White Paper: Analysis of truck driver age demographics across two decades—2014* (Report). Atlanta, GA: American Transportation Research Institute.

Shutko, J., Mayer, K., Laansoo, E., & Tijerina, L. (2009). *Driver workload effects of cell phone, music player, and text messaging tasks with the sync-voice interface versus the devices' handheld visual-manual interfaces* (SAE paper 2009-01-0768). Warrendale, PA: Society of Automotive Engineers.

Shutko, J., & Tijerina, L. (2011). Ford's approach to managing driver attention: SYNC and MyFord Touch. *Ergonomics in Design*, *19*(4), 13–16.

Sloss, D., & Green, P. (2000). *National automotive center 21st century truck (21T) dual use safety focus* (SAE Paper 2000-01-3426). Warrendale, PA: Society of Automotive Engineers.

Society of Automotive Engineers. (2002). *Calculation of the time to complete in-vehicle navigation and route guidance tasks* (SAE recommended practice J2365). Warrendale, PA: Author.

Society of Automotive Engineers. (2004a). *Navigation and route guidance function accessibility while driving* (SAE recommended practice 2364). Warrendale, PA: Author.

Society of Automotive Engineers. (2004b). *Rationale document for SAE J2364* (SAE information report J2678). Warrendale, PA: Author.

Society of Automotive Engineers. (2010a). *Definitions of driving performance measures and statistics* (SAE recommended practice J2409, draft). Warrendale, PA: Author.

Society of Automotive Engineers. (2014). *Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems* (SAE Information Report J3016). Warrendale, PA: Society of Automotive Engineers.

Society of Automotive Engineers. (2015). *Operational Definitions of Driving Performance Measures and Statistics* (SAE Recommended Practice J2944), Warrendale, PA: Society of Automotive Engineers.

Steinfeld, A., Manes, D., Green, P., & Hunter, D. (1996). *Destination entry and retrieval with the Ali-Scout navigation system* (Report No. UMTRI–96-30, also released as EECS-ITS LAB FT97-077). Ann Arbor, MI: University of Michigan, Transportation Research Institute.

Steinfeld, A., & Tan, H. S. (2000). Development of a driver assist interface for snowplows using iterative design. *Transportation Human Factors*, *2*(3), 247–264.

Stevens, A., Quimby, A., Board, A., Kersloot, T., & Burns, P. (2004). *Design guidelines for safety of in-vehicle information systems* (Technical Report). Crowthorne, UK: TRL Limited.

Swanson, A. (2016, January 4). How to predict the future better than anyone else. *Washington Post.* Retrieved from https://www.washingtonpost.com/news/wonk/wp/2016/01/04/how-to-predict-the-future-better-than-everybody-else/

Teo, L. & John, B. E. (2008). Towards predicting user interaction with CogTool-Explorer. In *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting* (pp. 950–954). Santa Monica, CA: Human Factors and Ergonomics Society.

Tetlock, P. E. and Gardner, D. (2015). *Superforecasting: The art and science of prediction*. New York, NY: Crown.

Thrun, S., Montemerlo, M., Dahlkamp, H., Stavens, D., Aron, A., Diebel, J.,…Lau, K. (2006). Stanley: The robot that won the DARPA Grand Challenge. *Journal of Field Robotics*, *23*(9), 661–692.

Tijerina, L., Angell, L., Austria, A., Tan, A., & Kochhar, D. (2003). *Driver workload metrics literature review*. Washington, DC, U.S. Department of Transportation, National Highway Traffic Safety Administration.

Transportation Research Board (n.d.) SHRP 2. Retrieved from http://www.trb.org/Strategic HighwayResearchProgram2SHRP2/Blank2.aspx

Trimble, T. E., Bishop, R., Morgan, J. F., & Blanco, M. (2014). *Human factors evaluation of level 2 and level 3 automated driving concepts: Past research, state of automation technology, and emerging system concepts* (Report No. DOT HS 812 043). Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.

Tsimhoni, O., Smith, D., Green, P. (2004). Address entry while driving: Speech recognition versus a touch-screen keyboard, *Human Factors*, *46*(6), 600–610.

Tsimhoni, O., Winter, U., & Grost, T. (2009, August). *Cultural considerations for the design of automotive speech applications*. Paper presented at the IEA Seventeenth World Congress on Ergonomics, Beijing, China.

UK Department for Transport. (n.d.). *Personal travel* (National Travel Survey). Retrieved from http://webarchive.nationalarchives.gov.uk/20100519172108/http:/www.dft.gov.uk/pgr/statistics/datatablespublications/personal/

U.S. Department of Transportation. (2008). *National motor vehicle crash causation survey* (Report No. DOT HS 811 059). Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.

U.S. Department of Transportation. (2013). *Visual-manual NHTSA driver distraction guidelines for in-vehicle electronic devices*. (Notice, 78(81), Part II, Docket NHTSA-2010-0053). Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.

U.S. Department of Transportation. (2015). *Traffic safety facts 2014* (Report No. DOT HS 812261). Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.

U.S. Department of Transportation. (2016a). *Distracted driving 2014 (traffic safety facts research note)*. (Report No. DOT HS 812 260). Washington, DC: U.S. Department of Transportation, National Highway Traffic Safety Administration.

U.S. Department of Transportation. (2016b). *Highway statistics 2015*. Washington, DC: U.S. Department of Transportation, Federal Highway Administration, Office of Highway Policy Information. Retrieved from https://www.fhwa.dot.gov/policyinformation/statistics/2015/

U.S. Vehicle Dependability Study VDS. (2016, February 24). *J. D. Power*. Retrieved from http://www.jdpower.com/press-releases/2016-us-vehicle-dependability-study-vds

U.S. Department of Transportation. (2017). *Distracted driving*. Washington, DC: Author. Retrieved from https://www.nhtsa.gov/risky-driving/distracted-driving

Violanti, J. M., & Marshall, J. R. (1996). Cellular phones and traffic accidents: An epidemiological approach. *Accident Analysis and Prevention*, *28*, 265–270.

Wang, P., Winter, U., & Grost, T. (2015). Cross cultural comparison of users' barge-in with the in-vehicle speech system. In *Fourth International Conference on Design, User Experience, and Usability: Interactive Experience Design (DUXU 2015)* (pp. 529–540). Cham, Switzerland: Springer.

Wang, H., Zhou, L. B., & Ying, Y. (2010). A novel approach for real time eye state detection in fatigue awareness system. In *Proceedings of the 2010 IEEE Conference on Robotics, Automation and Mechatronics* (pp. 528–532). New York, NY: IEEE.

Winter, U., Grost, T. J., & Tsimhoni, O. (2010). Language pattern analysis for automotive natural language speech applications. In *Proceedings of the Second International Conference on Automotive User Interfaces and Interactive Vehicular Applications* (pp. 31–41). New York, NY: ACM.

Winter, U., Tsimhoni, O., & Grost, T. (2011, June). *Identifying cultural aspects in use of in-vehicle speech applications.* Paper presented at the Afeka AVIOS Speech Processing Conference, Tel Aviv, Israel.

World Health Organization. (2013). *Global status report on road safety*, Geneva, Switzerland: Author.

Xiong, H., Bao, S., & Sayer, J. (2014). Factors affecting drivers' cell phone use behavior: Implications from a naturalistic study. *Transportation Research Record*, *2434*, 72–79.

Young, R. (2012). Event detection: the second dimension of driver performance for visual-manual tasks. *SAE International Journal of Passenger Cars-Electronic and Electrical Systems*, *5*(2012-01-0964), 297–316.

Young, R. (2015). *Revised odds ratio estimates of secondary tasks: A re-analysis of the 100-Car naturalistic driving study data* (SAE paper 2015-01-1387). Warrnedale, PA: Society of Automotive Engineers.

Ziomek, J., Tedesco, L., & Coughlin, T. (2013). My car, my way: Why not? I paid for it! *IEEE Consumer Electronics Magazine*, *2*(3), 25–29.

41

Improving Ad Interfaces with Eye Tracking

Michel Wedel

Gaze Recording for Ad Interface Design

As digital devices continue to pervade nearly every aspect of consumers' lives, consumers' daily experiences are increasingly shaped by the user interfaces (UI) of these devices. Digital screens provide the entry to information for work, travel, gaming, information, entertainment, social interactions, and shopping. Digital ads have become an integral part of the user interface, and in some cases commercial messages are even difficult to distinguish from online content. Digital advertising, a staple of business practice today, is a cost-effective way to place relevant, targeted ad messages.

But, given low click-though rates, the testing, evaluation and improvement of online ad interfaces is critical. Fitts, Jones, and Milton (1950) conducted the first usability study with eye tracking, which provided methods and findings that have become central in user interface research. Eye tracking has now become one of the methods of choice to evaluate UI design, digital product displays and digital advertising. Low- and mid-level features, including color and edges, and the size, number, location and configuration of ad elements are a key focus for enhancing ad interfaces through eye tracking research. Natural user interfaces (NUI) are being developed that improve ad effectiveness by rendering advertising more immersive, especially when it is combined with virtual reality and augmented reality technology. New forms of advertising may allow users to interact with them through NUIs that use speech, gesture, eye, and face recognition. Eye movement recording is likely to become an even more integral part of the NUI in the near future, allowing users to interact with ads by looking at them. Against this backdrop, the present chapter discusses eye tracking research in marketing, and discusses its future outlook.

There has been a rapid growth in commercial applications of eye tracking since 2000. The purpose of these studies is to facilitate perception, exploration, search and decision making of commercial products. Firms use eye-movement research in the design of digital advertising, online shelf layout, websites, apps, e-mails, and many more. Demand from marketing practitioners is increasing and, in response, providers of eye-tracking research conduct many hundreds of studies each year. Several global market research companies such as Perception Research Services

The Wiley Handbook of Human Computer Interaction, Volume 2, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

(PRS) (www.prs-invivo.com/) offer these services, and many large companies deploy the technology in their own dedicated research departments. Impetus for the continued growth of eye tracking comes from technological innovations and sharp declines in the costs of eye-tracking methods. Earlier, the diffusion of commercial applications of eye tracking was hampered by the high levels of specialization and expertise that were required, and by erroneous beliefs about limitations on the role of visual attention in information processing.

Recently, however, new developments in gaze recording technology have expanded the opportunities for widespread application of eye movement recording. Modern eye-tracking equipment has been miniaturized and most commercial eye trackers are portable. This enables eye tracking not only on desktop and laptop computers but on tablets and smart phones, in stores, games, vending machines, and digital billboards. Software solutions now make it possible to record viewers' gaze via webcams (http://thirtysixthspan.com/openEyes/; http://www.inference.phy.cam.ac.uk/opengazer/) (see Chapter 21 in this handbook). As front-facing cameras have become standard on computers and mobile devices, this creates enormous opportunities for remote eye-tracking studies on large panels of respondents. Participants can be tracked sitting in front of their desktop, laptop or tablet at home, at work, or anywhere else, as long as they have a webcam and Internet connection. However, eye tracking through webcam-based corneal imaging currently still lags somewhat behind in accuracy and quality, and data collection in naturalistic settings may sometimes be hampered by calibration issues. But large sample sizes and more natural conditions compensate for these limitations. Companies such as Sticky (http://www.gazehawk.com/; https://www.sticky.ai/; http://nviso.ch/) use webcams to record remotely not only gaze but also emotions through facial recognition.

In this chapter we review eye movement recording and its applications in marketing, with a focus on findings on the effect of image characteristics that can gainfully be manipulated in the design of visual stimuli and user interfaces. Much of the research on advertising in marketing has yielded findings on low- and mid-level features that are pertinent for online advertising as well. We begin with an exposition of eye movements and eye movement measurement, focusing on video recording. We then provide a brief history of eye-tracking research in marketing, and review general findings on the effect of ad image features on users' gaze. A prior review was provided by Wedel and Pieters (2008). We end with an outlook on the future of eye movement data and its use in NUIs for advertising.

Eye Movements and Eye-Movement Recording

Eye movements

Researchers such as Buswell (1935) and Yarbus (1967) were among the first to reveal that eye movements can provide detailed insights into visual perception processes of consumers. Since then, perception, exploration, search and decision-making processes have been extensively studied using observable measures derived from eye-movement recordings. It has become clear that eyetracking can provide insights into these processes that cannot be obtained by survey-type self-reports of present or past visual attention (Aribarg, Pieters, & Wedel, 2010; Russo, 1978).

We believe that we see the complete visual field at a high resolution in full detail at all times. This is due to the fact that if we look at certain aspects of a digital screen, whatever we focus our attention on is clearly visible. Everytime we move our eyes and pay attention to something else, we see it in full detail again. This gives the impression of "complete vision": that the detail of what is in our visual field can be perceived completely and immediately. This impression is incorrect. Introspection does not tell us precisely how we move our eyes, and what we see and what we miss. Eytracking is needed to establish what people look at and what they see and what not.

Hering (1879) and Javal (1879) were the first to observe that what was believed to be smooth movements of the eyes during reading in reality consist of bursts of discrete movements: fixations and saccades. The visual acuity provided by the retina of the human eye rapidly falls off with increased distance from the center, which is called the fovea and is the most senstitive part of the retina directly opposite the pupil. Because of this, we see less than 1% of what is in our visual field perfectly clearly at any particular moment. Therefore, in order to attend to specific object or location, we have to move our eyes, which makes it of interest to study eye movements as indicators of what information we acquire. There are six muscles that control movement of each eye, plus one muscle that controls movement of the eyelid (*levator palpebrae*), and one that controls the opening of the pupil (*dilator pupillae*). Four of these muscles control the movement of the eye in the up (*superior rectus*), down (*inferior rectus*), left and right (*lateral* and *medial rectus*) directions. Up-down movements are controled by the same muscle in each eye. Horizontal gaze shifts involve one eye moving towards the lateral side of the head and the other moving towards the medial side. Intermediate eye movement directions are controlled by simultaneous actions of multiple muscles. The other two muscles counteract head movement to maintain a stable gaze direction, and pull the eye in downward/medial (*superior oblique*), respectively upward/lateral (*inferior oblique*) directions. The eye-movement muscles are even slightly active when the eye is still. This tonic activity serves to keep the eye in place and maintains vision by stimulating the retina. Muscle activity is coordinated to make the eyes move together in this manner. Eye movements are both under voluntary and involuntary control but are mostly executed without conscious effort. As a consequence, eye movements are both precise and fast.

Saccades are quick jumps of the eyes that help to redirect the line of sight to a new location in the visual field and project it onto the fovea. A saccade takes 20–100 ms and reaches speeds of up to 1000 degrees/s. We make around 170,000 saccades per day. Saccades are often automatically driven by something that stands out in terms of the configuration of basic visual features, such as contrast or movement, in the periphery of vision. Yet, it can also be consciously controled, if we intentionally redirect our line of sight with the purpose of inspecting a location or object further. During a saccade, vision is suppressed. As a consequence, each day while we are awake, we do not see anything for as much as one-and-a-half hours!

An eye fixation typically lasts around 200–500 ms, depending on the task. During this time, a small region in the visual field is projected onto the fovea, which enables it to be processed in detail. The fovea accounts for about 2 degrees of visual angle of the visual field: about a thumbnail at arms length. Even during a fixation, the eye is never completely still and moves slightly. These movements include drift, tremor and microsaccades. Microsaccades are tiny movements, about 0.2 degrees of visual angle. Further, vergence eye movements are movements where both eyes turn inward, in order to keep an object that moves towards or from us in focus (or vice versa), and smooth pursuits are eye movements that enable us to follow objects moving across the visual field, with speeds of up to 100 degrees per second.

Eye-Movement Recording

Eye-tracking devices were invented around 1900. Huey (1898) developed an eye tracker that used a lever attached to a cup placed on the eye with a hole for the pupil, which moved a pen across the surface of a smoked drum to record movements of the eye. To overcome the mechanical limitations of this method, Orschansky (1899) recorded light that was reflected off a mirror attached to the eye cup. An even more accurate method was to record light reflected off the surface of the eye itself. The first eye tracking device utilizing this principle was developed by Dodge (1900), called the "falling-plate" camera. The device involved a holder for printed text, to be read by the participant. On a photographic plate, a camera recorded the reflection of sunlight that bounced off a white piece of cardboard on the front of the eye. Bicycle pumps were used to ensure smooth descent of the "falling plate," and a swinging pendulum caused a regular pattern on the falling plate, creating a time record. Knitting needles were used to hold pieces of white paper in place that served as fixation points. This device produced the first published trace of eye movements and was later manufactured and distributed professionally by Spindler, Hoyer, and Göttingen (Wade, 2010).

Today, eye trackers measure eye movements in one of three principal ways: (a) measurement of the movement of a special contact lens attached to the eye, (b) measurement of electric potentials generated by the eye movement muscles using electrodes (electro-oculography), and (c) video-based tracking of reflections of light on the eye (video-oculography). Most commercial eye-tracking methods and devices fall in the latter category. This method has become popular because it is noninvasive and relatively inexpensive. There are two main types of video-oculography: those that use infrared (IR) light emitted from a source, and those that do not.

In IR-based video-oculography, infrared light is emitted from a light source and reflected off the eye, thereby creating a reflection on the cornea, the outer layer of the eye. Typical commercial eye-tracking devices measure that reflection on the cornea of one or two eyes (more precise eye trackers use two or more of these so-called Purkinje reflections on each eye). The distance between the center of the pupil and the corneal reflection changes with eye movements, but remains relatively constant with normal head movements. Therefore, these modern eye-tracking devices allow for fairly wide head movements during recording, and thus more unobtrusive measurement. The point of regard (POR), the location where the eyes focus on, is determined from the angle and distance of the corneal reflection from the pupil center, after a calibration task. Bright-pupil and dark-pupil methods differ in the location of the IR light source relative to the eye. If IR light enters the eye along the optical path, then it reflects off the retina back into the recording camera creating a bright pupil. If the IR source is at an angle to the optical path, then the pupil appears dark because the reflection of the light by the retina is directed away from the camera. Bright-pupil tracking allows for more robust eye tracking and is less prone to interference and variations in light, but cannot be used outdoors. Modern eye trackers switch seamlessly between bright- and dark- pupil recording, which increases their accuracy. Commercial IR-based eye-tracking typically has a temporal resolution of about 50 Hz (20 ms) (or multiples of this), and spatial resolution of 0.5°, which is sufficiently accurate for commercial and many academic applications. But, the temporal resolution needed for fundamental research can be much higher, and some eye trackers may provide up to 2000 Hz resolution (in which case the head of the participant needs to be fixed). Eye-tracking devices record the *x*- and *y*- coordinates of the POR of both eyes at that specific frequency, and the continuous trace of the POR has to be processed further before analysis by identifying fixations, saccades, smooth pursuits, and sometimes microsaccades as well.

Various companies, including Tobii and SMI (www.smivision.com; www.sr-research.com; www.Tobii.com) offer table-mounted and portable and lightweight eye-tracking devices. Miniature head-mounted eye trackers embedded in glasses are rapidly gaining ground in applied research. The latter are particularly useful when eye recording during unrestricted movement of people is required, such as during (digital or virtual) shopping tasks, or exposure to multiple digital screens. The comparatively low costs of these new generations of eye-tracking systems, short calibration times, unobtrusive measurement, the possibility to track consumers with eyeglasses and contact lenses, and improved processing software have all contributed to the exponential growth of eye-movement applications in practice, gaze measurement in natural exposure conditions, and applications for theory development and testing in academic research.

A more recent development involves IR-free eye tracking through geometric analysis of corneal images obtained via front facing cameras on mobile devices and desktop and laptop computers. One of the leading companies providing webcam-based eye tracking is Sticky (https://www.sticky.ai/). Innovations in IR-free eye tracking are due to the initial developments of researchers such as Pelz et al. (2000) and Li and Parkhurst (2005). It consists of software solutions to analyze the video from front facing webcams. Chapter 21 in this handbook provides an extensive description of methods and applications. Open-source software is available (http://thirtysixthspan.com/openEyes/). The software recognizes specific features of each eye from the image, including the eyelids, pupil and its contour, the iris, the limbus (the border of the iris), and the corners of the eye. Blinks are used to locate the position of the eye. Algorithms are then applied to identify the center of the pupil and the gaze direction. These algorithms are based on the principle that the contour of the pupil or iris, captured by the webcam and projected on the two-dimensional video image, is an ellipse. Because in three dimensions the contour is a circle, the orientation of the ellipse can be used to mathematically reconstruct the original circle in three dimensions. This, along with other calibration points on the eye, allows the center of the pupil and orientation of the gaze to be calculated. The video image is typically captured at a frequency of 10 Hz (100 ms), which provides a somewhat lower temporal precision than typical IR-based eye tracking. A comparison of the results of IR-based and IR-free eye tracking (Burton, Albert, & Flynn, 2014) shows that IR eye tracking is more accurate for smaller regions of interest (around 1% of the screen or less), particularly if they are located in the periphery of the computer screen. Because webcam-based eye tracking misses some of the precise locations of the eye, it tends to underestimate dwell time, on average by as much as 50% for these smaller images. For larger images (about 5% of the screen or more); however, the webcam technology achieved accuracy comparable to IR-eye tracking, especially in capturing the percentage of participants who fixated on the region of interest. Dwell times were underestimated by about 25% for these images. The authors conclude that for research that focuses on reasonably sized regions of interest, not located too far in the periphery of the screen, webcam based eye tracking is a viable technique, especially for the purpose of identifying the percentage noted. But, the downward bias of measures provided by IR-free eye tracking is noticeable and needs to be kept in mind when using this approach.

The advantages of IR-free webcam based eye tracking are many. First, it brings down the cost of eye tracking by an order of magnitude. Second, to offset the lower spatial and temporal precision, this type of eye tracking can be done on very large samples of respondents, reducing the measurement variance. Third, IR-free methods allow eye tracking to be done in natural settings, at home, at work, or any other location where respondents are in front of a desktop or laptop computer—and in the future probably tablets and smartphones as well. Fourth, webcam-based eye tracking makes research across dispersed geographic locations, especially crosscultural and crossnational studies, feasible, and renders it more standardized and more cost effective. Fifth, an image of the scene that the subject in question looks at can often be obtained from the corneal image as well, as it is reflected on the cornea. Sixth, this method holds great promise for gaze control in NUIs.

Computing fixations and saccades

Eye-movement recording produces a sequence of locations where the eyes focus on, the points of regard, typically consisting of a time stamp and an *x*- and *y*- coordinate for each eye. From this raw trace of the POR, eye fixations and saccades need to be identified as the key eye movement measures. Sequences of eye fixations, called scan paths (Noton & Stark, 1971), are the critical measures used to understand in visual behavior. Various algorithms have been proposed to compute eye fixations from the recording of the POR that the eye-tracking equipment provides. Most algorithms commonly used in eye-movement research identify fixations, and thus saccades between them, based on prespecified thresholds of velocity, distance, duration, angle, or acceleration of the POR. Dispersion-based methods are most frequently used because their ease of implementation. Recently, however, velocity-based algorithms have gained popularity because they are more transparent and more accurate in identifying the precise beginning and end of saccades. Van der Lans, Wedel, and Pieters (2011) developed the "binocular-individual-threshold" (BIT) algorithm, which is an open source, fully automatic parameter-free fixation-identification algorithm that identifies task- and individual-specific velocity thresholds by optimally exploiting the statistical properties of the eye-movement data across both eyes. Anomalies caused by blinks and tear fluid are filtered out of the raw trace of the POR. The BIT algorithm defines consecutive sequences of samples of the POR and marks them as fixations or saccades. Fixation counts, fixation durations, gaze durations, gaze selection, time to first fixation, and fixation transitions on regions of interest (ROI) are among the measures most often analyzed in eye movement research. Blinks, head movements and the diameter of the pupil are obtained as a corollary of video-based eye tracking.

A Brief History of Eye-Tracking Research in Marketing

In the early 1900s, Nixon (1924) and Poffenberger (1925) first applied eye-movement research to determine consumers' attention patterns to magazine and newspaper advertisements. Nixon (1924) manually recorded eye movements of consumers who paged through a magazine with print ads, while hiding himself in a box behind a curtain. Karslake (1940) used an eye camera to collect eye-movement data on advertisements appearing in a newspaper, which yielded higher efficiency and better accuracy. Yet, in marketing, there followed a period of relative paucity in research activity, possibly caused by the popular but erroneous view that attention is only the first stage towards higher order cognitive processes. This view was advocated through hierarchical information processing models such as AIDA (Starch, 1923). New impetus for the use of eye tracking came from Russo's article in 1978, titled: "Eye-fixations can save the world," in which he argued for the use of eye movements to study information processing, focusing on consumer decision processes. Russo compared eye tracking with other cognitive process tracing methods (information display boards, input-output analysis, and verbal protocols) on a range of criteria. He concluded that eye-tracking methods offers many advantages of validity, unobtrusiveness, ease-of-use and cost, not offered by these other methods. Once it became established that eye movements are tightly coupled with visual attention and that they can be used to infer information acquisition and higher cognitive processes, there was a revival of interest in eye tracking that started in the 1970s. Much of this involved descriptive studies that documented attention to such stimuli as Yellow Pages ads, nutrition labels, and alcohol and cigarette warnings (see Wedel & Pieters, 2008 for a review). The 1990s then saw a surge of interest, in part driven by advances in commercial IR-based recording technology. Applications of eye tracking have appeared in studies involving in-store choice decisions and shelf search, print advertising, TV commercials, e-commerce, labeling and educational messages, and branding.

From then on, the increasing reliably, robustness and unobtrusiveness of the measurement of visual attention with eye tracking, and the recognition of its central role in consumer behavior with respect to visual marketing stimuli led to more and more applications, and testing of theories from fundamental attention research. This enabled the evaluation of long-standing beliefs in marketing practice and it facilitated predictions of the effectiveness of commercial visual stimuli, leading to many new insights. Marketing provides rich natural contexts in which consumers engage with many complex stimuli (websites, apps, catalogs, shelves, packages, ads, video, commercials), through a variety of tasks (exploration, search, choice). Since around 2000, the Bayesian framework for developing and estimating statistical models of visual attention has played a crucial role in eye-tracking research to disentangle multiple underlying cognitive processes from eye-movement data (Wedel & Pieters, 2000). This research has yielded a range of fundamental insights on the effects of low- and mid-level visual features in ad images. These are reviewed next.

The Effect of Visual Features

We review eye-movement research in marketing, focusing on the effect of low-level visual features, including size, position, color, edges, but also mid-level characteristics such as originality, complexity, and clutter, in ad design. Many of these features are pertinent to web usability and online advertising, as websites and online ads consist of mixtures of textual and pictorial information and contain multiple objects that compete for attention.

Effects of size

Much is known about the effects of size of key regions of interest (ROI) in ads. Size clearly matters in capturing attention and bigger is better, in general. However, for complex stimuli such as ads and websites, with both text and pictorial information, the effects are more intricate and different from what is often tacitly assumed. Several measures of visual attention, including visual selection, fixation frequency, and total gaze duration, are strongly associated with the size of visual stimuli. For ads as a whole, gaze duration on an ad increases with 81% for every 100% increase in their size (Pieters & Wedel, 2004). For smaller feature ads (Free Standing Inserts) gaze duration increases by a smaller amount: 22% for a 100% increase in size (Pieters, Wedel & Zhang, 2007). This smaller effect for features ads may be caused by higher competitive clutter, because multiple of these small feature ads are usually placed on a single ad display page. Thus, it may be more difficult to use size to stand out in cluttered visual environments. Three elements of ads have received much research attention, the text, the picture and the brand. These are discussed next.

Text The increase in gaze on the text in an ad is almost proportional to its size: a 100% increase in surface size of the text leads to a 90% increase in gaze (Pieters & Wedel, 2004). This is likely due to the time it takes to read text. Ads with larger text elements receive longer gaze (Pieters & Wedel, 2004), which is probably due to reading. Rayner, Rotello, Stewart, Keir, and Duffy (2001) therefore conclude that text is more important in designing ads than is generally assumed. Yet, for feature advertisements, Pieters et al. (2007) showed that text should be decreased in size with about 20% relative to its current value. Because clutter for these types of ads is very high, these results indicate that as clutter increases, text is inspected less.

Pieters and Wedel (2007) showed that the goals consumers have while looking at ads (for example, exploration, learning about the brand or ad) affect gaze on text and pictorial elements. Text was looked at longer when consumers wanted to learn about the ad or the brand. These findings explain the long gaze on the text found by Rayner et al. (2001), who used a brand learning task. Thus, it appears that consumers' beliefs that the text in an ad contains most information—although not necessarily correct—influences their eye movements. From these studies, it is not clear to what extent the size effect for text is associated with an increase in font size, or an increase in the number of words. If the latter would be the case, then the interpretation of Rayner et al. is valid and advertisers aiming to maximize attention to the entire ad should devote more space to text, especially for targeting those consumers that have the goal to learn about the brand. However, qualitative inspection of the scan path on the text of ads in many eye-tracking studies reveals that text is usually not read in its entirety: consumers read either a fraction, and / or sample the text for keywords; but, the proportion of the text that is read increases with the involvement of the viewer (Rosbergen, Pieters, & Wedel, 1997).

Most of the studies to date have failed to find a significant transfer of gaze from text to the picture (e.g., Pieters & Wedel, 2004; Rayner et al., 2001). Instead, gaze transfers in the other direction: from the pictorial to the text. Also, text does not seem to contribute to brand recognition (Wedel & Pieters, 2000). These findings cast some doubt on the central role of text on attention and memory, although it might be important for comprehension. The research by Stewart, Pickering and Sturt (2004), shows that longer gaze on text may result from semantic incongruence or complexity.

Pieters, Wedel and Batra (2010), however, do not find an effect of complexity of ads on gaze duration on the text in print ads. These somewhat inconsistent findings point to the need for more research on the role of the text in multimodal stimuli such as ads and websites. The literature on eye movements in the context of reading (e.g., Rayner, 1998) provides limited guidance because advertisements are usually multimode messages with pictorials, text and symbols. As part of the text, the specific role of the headline is under-researched as well. The headline stands out due to its size, large font type, and central position in the ad. Generally, the idea in advertising is that the headline attracts gaze early, and needs to draw consumers into spending more time with the ad. Few studies, however, have specifically looked into gaze capture by and transfer from the headline to the other elements. Rosbergen, Pieters and Wedel (1997) found that the headline receives longer dwell times than the body text. In addition, while virtually all consumers in that study inspected the headline, this was not the case for the text. Further, Leven's (1991) study showed that the eyes tend to quickly go to the upper left corner of an ad, where usually the headline is located. Thus, expectations on the layout of ads may have strong effects on gaze patterns on the headline and body text.

Picture The pictorial tends to capture gaze regardless of its surface size. Dwell times on pictorial elements, including a brand logo, have been found to increase with increasing size but at a lesser rate than is the case for text: a 100% increase leading to a 30% increase in gaze. The effects of increasing the size of the pictorial on dwell times are relatively small and counter to common beliefs: increasing its size has no effect on gaze on the ad as a whole (Pieters & Wedel, 2004). Although picture size has been found to affect memory, the effects are modest in size (Wedel & Pieters, 2000). Based on current findings, advertisers and web designers would be ill advised to maximize the size of the pictorial elements in an effort to maximize gaze on the ad. It has been found, however, that the pictorial elements in ads are critical for ad gist perception. Consumers are able to recognize an ad and the product category it promotes within as little as 100 ms, if the ad has a typical layout (Pieters & Wedel, 2012). The picture plays an important role in this.

For feature ads the effect of the size of the pictorial element is larger than for print ads. Pieters et al. (2007) showed that among all elements of the ad the size of the picture has the largest effect on dwell time. An explanation for this is again the high competitive clutter for feature ads. Under such high competitive clutter, the picture elements may capture more gaze when it is larger. Nevertheless, Pieters et al. (2007) reported that in current feature advertisements, the pictorial is still too large, and can be decreased in size with close to 40% to achieve an optimal composition of the ad display page that would capture maximum gaze. In their research on the effect of goals, Pieters and Wedel (2007) showed that when consumers wanted to learn about brand they looked less at the pictorial, reflecting their belief that the picture contains less information on the brand than the text.

Many aspects of the pictorial other than its size are not yet very well understood. In particular, aspects of the content of the pictorial, such as the presence of specific objects and their complexity are underresearched. Wedel and Pieters (2015), for example, recently discovered that a central identifying object displayed in the picture is critical for accurate gist perception of ads, especially in blurred exposures of as little as one tenth of a second. This is important, because if, in the very first fixation on an ad, consumers are able to determine that they are looking at an ad and can identify what is being advertised, this may prompt further gaze and deeper processing of the information contained in the ad.

Brand Studies to date point to a crucial role of the brand or product logo in the ad. The effect of the brand size is similar to that of the pictorial but gaze captured by the brand element transfers to the pictorial and text more than the other way around. Thus, the brand plays a key role in routing gaze across ads. In addition, gaze on the brand carries over to memory for the brand more strongly than those of any of the other ad elements (Wedel & Pieters, 2000). Thus, capturing gaze on the brand by increasing its size is likely to transfer to the key message contained in the picture and text, and improve memory for the brand (Pieters & Wedel, 2004). The results for feature ads (Pieters et al., 2007) are consistent with this: the size of the brand has a strong positive effect on gaze, and if current feature ads were to be optimized, the size of the brand should increase by about 75%. Relatedly, Chandon, Hutchinson, Bradlow and Young (2009) found that more space devoted to a brand on the shelf increases dwell times.

There is relatively little research into the effects of sizes of other elements or ROIs than the ones discussed above. Pieters et al. (2007) show that the size of ROIs that contain price and / or promotional information has positive effects on gaze on feature ads. In feature advertising these elements provide key information. If ads were designed to be optimal, price should be about 60% larger, and promotions about 10% larger than they currently are in practice. Some of the results of size effects on gaze on print and feature ads may be generalized to other marketing stimuli, such as catalog pages, shelves, coupons, websites, billboards, and banner ads, and the study by Dreze and Hussherr (2003) affords useful initial insights into the latter.

The context and competitive configuration may exert a major influence on dwell times as well, which interacts with some of the effects of size discussed above. As yet, little is known on the effects of context. The work by Janiszewski (1998) first revealed that competition plays a major role and may moderate the effects of sizes of elements and objects in ads. Wedel and Pieters (2000) establish the effects of context variables in magazines, left-right page location and ad serial position (page number), but more as control variables in a study that primarily deals with size effects on attention and memory. Consistent with these findings, Pieters and Wedel (2004) also found negative effects of page number on dwell time on the ad as a whole, which may point to the negative effects of competition of ads seen before the focal ad, which increases with the page number.

Position

Position effects on eye movements have been demonstrated in a number of studies. First, several studies found strong effects of top / center spatial location on eye movements, sometimes called the "center-of-gravity" effect (Buswell, 1935). D'Ydewalle and Tamsin (1993) found that a central position of billboards on a TV screen during a soccer match captures more gaze. Leven (1991) found short times to the first fixation on the top-left position in ads. Dreze and Hussherr (2003) found positive effects of a central location of banner ads on Web pages. Studying centrally placed banner ads, Josephson (2005) finds strong positive effects of a placement on the top, versus the bottom of a webpage. Liechty, Pieters and Wedel (2003) (see also Wedel, Pieters and Liechty, 2008) reported that, while browsing ads, consumers nearly always start with local exploration in the center of the page, but Chandon et al. (2009) report that top- and middle-shelf positions receive longer gaze. The center may be an optimal location for early information processing, or it may be a convenient location from which to start exploration of an ad, a screen, a shelf, or a website. The center may not only receive more eye movements during the initial stages, but also during the final stages of processing. Atalay, Bodur and Rasolofoarison (2012) found progressively increasing gaze on the central option on a shelf, prior to a decision. Brasel and Gips (2008) found that during fast forwarding, brands located in the center of the screen receive more gaze, the reason being that when fast forwarding to the ad, the viewer needs to pay close attention to determine the start of the show she intends to watch.

Second, some research has demonstrated systematic left-right spatial eye-movement tendencies on a variety of visual stimuli. Van der Lans, Pieters and Wedel (2008a) show that the scan path on a shelf demonstrates systematic left-right strategies, which suggests that eye movements are guided by the horizontal shelf organization. Similarly, Shi, Wedel and Pieters (2013) show that while making a choice on comparison websites, which display products in the rows and their attributes in the columns, eye movements showed a strong left-to-right tendency. However, consumers switched about once every second between horizontal and vertical eye movements. For ad exploration, systematic left-right eye movement tendencies have not been reported (see Wedel & Pieters, 2008), with the exception of eye movements made while reading the text or the headline. It seems that left-right eye movements are primarily prompted by the organization of the visual stimulus, such as a shelf, comparison website or text, rather than being an deep-rooted eye-movement behavior. More research is needed to support that claim, however.

Third, several studies have demonstrated that exploration of a complex scene is broken up into bursts of local explorations of the most informative regions, interspersed with global jumps between those regions. That is, across a range of tasks, including exploration, search and decision making, eye movements have been found to cluster in relatively small contiguous regions on the visual display. In several types of visual stimuli, these contiguous regions correspond with objects or elements of potential interest to the viewer. Liechty et al. (2003) investigated the spatial distribution of eye movements on ads, and distinguished between local (fixations on neighboring cells of a spatial grid) and global (fixations on nonneighboring cells) patterns. These authors concluded that the scan paths that they observed serve the purpose of breaking down the complex task of ad exploration into a smaller number of simpler tasks that involve exploration of local regions in the ad. Consumers switched between local and global exploration about five times during exposure to an ad: in the local state contiguous regions are explored through foveal vision, while in the global state the eyes are redirected through peripheral vision. There was no evidence of a pattern of initial global exploration followed by local exploration, which had been postulated previously. Finally, similar eye-movement patterns were observed during choices on comparison websites, by Shi et al. (2013). They found that fixations were concentrated on contiguous elements of the display, predispositions that persisted even when the row / column orientation of the website was changed. Consumers attended to about two to three attributes or products before switching. The generality of these clustered local fixation patterns across very different stimuli and tasks is striking, but more research is needed to support generalizeability and into the factors driving these patterns.

Low-level image characteristics

Relatively few studies have investigated the effect of low-level features, such as colors, edges, textures, and contrast, in visual exploration tasks. Rather than systematic inquiries, several studies include a few selected features, and mostly lack strong theoretical explanation of the effects found. The most extensive studies on the effects of low-level features on eye movements are those by van der Lans, Pieters and Wedel (2008a, b). They studied visual search for brands on shelves, and demonstrated the effects of color and edges on eye movements and search performance. They reveal that colors and edges have strong effects on both localization and identification of brands on shelves. Which colors are salient depends in part on the search goal and varies substantially across individuals. Colors that are diagnostic for the search target (i.e. the color blue in a package) attract most fixations. The results suggest that consumers use only one or two basic features at the same time. The results also show that consumers actively direct the focus of their gaze away from the edges and to the center of brands (confirming the center-of-gravity effect), where more diagnostic information is expected to be located. These bottom-up effects explained about two thirds of visual salience.

The study of Wedel and Pieters (2015) shows that color has an important buffer effect during gist perception. When exposures are brief (less than the duration of a fixation) and blurred (because they occur in the periphery, in motion or at a distance), the color of the central object in an ad helps people to capture the gist of the ad, but under nonblurred (but brief) exposures, color had little effect.

Global image characteristics

Research has also examined effects of the global characteristics of complete ad images on gaze. Pieters, Warlop, and Wedel (2002) investigated the effect that originality and familiarity have on consumers' eye fixations on the brand, text, and pictorial, and on memory for the brand. They looked at "garden-variety" originality: ads that were within the normal range of originality found in magazines and newspapers. The results indicated that more original advertisements receive more fixations on the brand, which improves brand memory. Importantly, the effect of originality was even stronger for ads that respondents rated as more familiar. Radach, Lemmer, Vorstius, Heller, and Radach (2003) showed that the specific type of originality used by mystery ads, by having pictures and text that are not related to the advertised product, receive more eye fixations on various ROIs.

Pieters, Wedel and Batra (2010) investigated yet another implementation of originality that is achieved by making ads more complex. They distinguished two types of complexity. Feature complexity occurs when the ad image contains dense visual features. It can be measured by the GIF-compressed file size of the ad image. Ads that have high feature complexity typically come across as cluttered. Design complexity involves ads that have an ostentatious creative design and contain many different objects and elements, of irregular arrangement and shapes. The study showed that feature complexity hurts gaze on the brand. It makes the brand more difficult to find among the cluttered features. Consequentially, the brand ROI receives fewer fixations. Design complexity, however, in line with the above findings on other forms of originality, helps gaze on both the pictorial and the advertisement as a whole. This is good news but a caveat is that gaze on the pictorial does not carry over to the other ad elements, text and especially the brand. For TV commercials, Teixeira, Wedel and Pieters (2010) find that the tendency to stop watching commercials is higher when its feature complexity is low or when it is high, as compared to intermediate levels of feature complexity (feature complexity of each frame in a commercial was measured as the file size of that frame).

Clutter

The study by Pieters et al. (2010) revealed that feature complexity, measured as the file size of the JPEG ad image, negatively affects gaze on the brand. This measure of feature complexity can be interpreted as low-level clutter—that is, clutter in terms of the basic perceptual features. Its effects on gaze were found to be negative.

High-level clutter is related to the arrangement of objects in an image. Feature ads provide an opportunity to study this type of visual clutter because multiple ads, each with their own visual layout, appear on a single ad display. Many websites have this visual arrangement as well. Pieters et al. (2007) propose two measures of this type of clutter: target distinctiveness (TD) and distracter heterogeneity (DH), both measured through the entropy of ad element size distributions. Target distinctiveness measures how different a feature ad is from all other ads on a page, DH how different the other feature ads are from each other. Thus, each ad on a page with multiple ads has its own TD and DH measures. Feature ads with a more distinctive design (higher TD) were more often selected through eye movements. The higher its DH the less frequently a target ad is selected. Once an ad is selected, the gaze duration of consumers is affected only by TD, DH no longer plays a role. These effects of competitive clutter on eye movements present a tradeoff that is important in designing ad displays with multiple ads: increasing the distinctiveness of a particular ad improves both its visual selection through eye movements and the duration of gaze, but simultaneously doing so for all ads in an ad display increases heterogeneity of distracters (TD) and thus hampers visual selection of all ads. The research reveals, amongst other things, that a systematic arrangement of the ads on the page, and an orderly distribution of elements such as the picture, text, brand, and price, helps to alleviate this problem to some extent. It is even better to optimize each ad display page formally, which the authors do. The study provides a starting point for studies into competitive context and its effects on eye movements in a other settings, such as online product display pages, screens with app logos, search result pages, and comparison websites.

Downstream effects

The possible downstream effects of impacting gaze through ad and interface design are vast. There is substantial evidence that even small improvements in eye gaze on ads and other commercial visual stimuli, such as catalogues, shelves, labels, and packaging designs affect memory (Wedel & Pieters, 2008). This persistent association supports the importance of eye tracking research in the design of ads and user interfaces. Further downstream effects of eye gaze on a wide array of behavioral measures have been identified, including critical positive effects on attitudes, preferences and intentions (Pieters & Warlop, 1999; Rosbergen, Pieters &Wedel, 1997), search performance (van der Lans, Pieters, & Wedel, 2008a, b), consideration-set formation (Chandon et al., 2009), and decision making (Pieters & Warlop, 1999, Stüttgen, Boatwright & Monroe, 2012). Indeed, the use of eye movements to better explain and predict consumer decision making is a rapidly growing area that attracts interest from academics in marketing, psychology and economics alike (Ashby, Johnson, Krajbich, & Wedel, 2016).

Zhang, Wedel and Pieters (2009) showed that gaze on feature ads predicts subsequent retail sales in a large-scale national sample, while controlling for other variables. Their study reveals that gaze duration mediates the effect of low-level ad characteristics (sizes, color and location) on sales, which implies that the effect of low-level characteristics on sales is fully accounted for by the eye gaze patterns that they cause. The study by Zhang, Wedel, and Pieters (2009) is important because it is the first to establish the relation between gaze and sales. Nevertheless, the feature advertising context of their study is specific and most conducive to finding this type of relationship, because these ads are designed to have instantaneous effects rather than the intended longer term effects of print and other display ads. The gaze sales relationship needs to be established across a wider range of visual marketing stimuli, as a priority for future research.

Discussion and Outlook

The power of analytics and big data to enhance marketing effectiveness is widely recognized. Yet, much eye-tracking research in practice, unfortunately, as yet stops short of presenting more than simple graphical displays of eye-movement patterns. These heat maps, gaze plots and so on, are useful depictions of where respondents look most and what they miss. They provide qualitatively interesting insights, and are easy to communicate to clients. But, as reports become saturated with heat maps derived from small samples of respondents with little predictive power, their novelty fades and interest wanes, and the prospects of eye-tracking research are threatened. The application of analytics to eye movement data is key to the survival and continued success of eye-tracking research in practice. Eye-tracking research based on large samples of respondents, which will be increasingly feasible with gaze recording through webcams, will generate big data sets that require substantive expertise, proper analytical techniques and computing power. Those will facilitate efficient and effective eye movement analysis for a wide range of commercially important and academically interesting stimuli, including sponsored search, video clips, news, movies and trailers, digital billboards, banner ads, product reviews, product comparison sites, tag clouds, games, and virtual and augmented reality. Big gaze data collected on large samples of participants and large samples of marketing stimuli, when analyzed with proper analytical tools, will facilitate generalizations of findings, and enable the analytical optimization of visual marketing for optimal downstream impact.

With the development of corneal imaging software that records where consumers look via webcams, access to online information through a wide range of electronic devices is becoming more and more flexible because of the use of gaze control (see Chapter 20 in this handbook for applications). Companies such as Tobii, Samsung, and Lenovo have integrated gaze control in laptops, smart-TV and smartphones (https://s3.eu-central-1.amazonaws.com/theeyetribe.com/theeyetribe.com/index.html), using eye tracking via front-facing cameras to provide users with an easy and efficient way to control applications. The user can scroll through pages by swiping them with her gaze, and make a selection or activating applications by fixating on them. As part of NUIs this technology can support critical processes on digital devices such as exploration, search and decision making, and provide a direct and intuitive way to interact with digital content. It may allow for automatic alerts if important information is overlooked, and may use color, contrast, and blurring to interactively direct exploration and search. Consumers can already interact with 3D images on their smartphones through their eyes, in some applications overlaid on real world scenes. Google has developed glasses that show a viewer augmented information based on what she is looking at and what not, even without her ever having to explicitly search for that information, (http://marketingland.com/pay-per-gaze-advertising-new-google-patent-may-reveal-plans-for-monetizing-google-glass-55714) and Samsung's patented smart contact lenses can overlay computer generated images on real world scenes and can be controlled by the users' eye movements and blinks (http://www.independent.co.uk/life-style/gadgets-and-tech/news/samsung-smart-contact-lenses-patent-a6971766.html). As other augmented reality applications continue to emerge, in areas such as gaming (Pokemon Go: http://www.pokemon.com/us/pokemon-video-games/pokemon-go/), travel (Nokia's City Lens: https://www.engadget.com/2012/09/11/nokia-reveals-new-city-lens-for-windows-phone-8/), retailing (Yihaodian/Walmart's virtual stores: http://www.cnbc.com/2015/04/24/retails-new-reality-four-ways-technology-can-boost-sales-commentary.html) and education (The Smithsonian: http://naturalhistory.si.edu/exhibits/bone-hall/), gaze and face control and recording will find more important applications.

For example, the new gaze-control technology is already rendering computer games more absorbing. Gamers can already interact with avatars by looking at them (http://waterloolabs.com/). Moreover, as a player moves her head, stares, blinks, or expresses an emotion with her face, the avatar mirrors the viewers' movements and emotions (http://www.redbull.com/cs/Satellite/en_INT/Game/Red-Bull-Formula-Face-021243076152177). Using webcams, a user's face, eye, head and facial movements can be recorded and analyzed as an integral part of the UI to render the avatar's movements and expressions more realistic and more in tune with the user, producing games that are more appealing and immersive (http://www.optitrack.com/). Already, brick-and-mortar stores mannequins (http://www.almax-italy.com/) can tell if a shopper is male or female, estimate her age, number of companions, record how much time she spent looking at the mannequin and its outfit, and capture keywords in conversation between shoppers. The technology can also be used to support learning. Interactive online and mobile education can benefit from head, face and eye movements of students, not only to improve the interface, but also to improve student focus and comprehension in real time. Explanations, definitions and translations may appear when eye movements indicate that comprehension is slow. Reading problems may be diagnosed, reading-skill levels may be classified automatically, and the interface may automatically adapt to the user's skill level. Selective blurring and focus of text may improve reading concentration and speed (http://text20.net/). Dynamic gaze cueing may direct students' gaze to the most relevant aspects of the material. Further, on e-readers, gaze can be used to create document summaries and adaptively personalize them to the reader's interests. Recommendation systems may in the future recommend new articles, texts or reviews, based on the content of what a user has looked at previously. The implementation of eye tracking on digital devices makes it possible for keyword and image search results to be displayed dynamically, using the user's moment-to-moment gaze on keywords or images as input.

Facial recognition through webcams may also improve the targeting of marketing effort. User interface design determines how efficiently consumers find information on products and services across multiple screens. Technology to charge advertisers based on the time consumers spend looking at their ads already exists (http://marketingland.com/pay-per-gaze-advertising-new-google-patent-may-reveal-plans-for-monetizing-google-glass-55714). Gaze-based rendering, 3D vision and augmented reality are beginning to play an important role in enhancing the online shopping experience. Consumers can upload their body measurements to an online retailer along with pictures, to virtually try on new apparel and see how it looks. By integrating advertising recommendation engines with face and gaze tracking on smart-TV (http://www.mirametrics.com/), advertisers can adaptively customize ad content dynamically based on viewers' prior gaze and emotional expressions. Interactive gaze cueing may be used to direct viewers to product placements, and augmented reality allows virtual products to be placed where a person looks. If a viewer focused on the brand in a product placement during a TV show, a gaze-contingent recommendation engine can show a personalized ad for that product during the subsequent commercial break. Attentive and interactive billboards, digital ads, digital out-of-home displays and digital point of sale devices (http://cornermedia.com/) can in the future adapt dynamically to the density of traffic or viewers, and to their gaze, facial expression, and head and body movements. Dynamic billboards already exist that present more content-rich ads if traffic slows down, and simpler typical ads that can be processed rapidly if it speeds up. Microsoft's Natural User Interface Advertising (NUads) on Xbox allows users to interface with commercials using speech and gestures, to post the ad on social media, or ask for more information such as a map of a local retailer (https://www.youtube.com/watch?v=RSk5DhxQHLo). Gaze and face recognition will in the future become an integral part of users' interface with NUads.

Eye and face tracking through front-facing cameras will soon be an integrated part of our everyday lives, and be incorporated in the UI's of desktop, laptop and tablet computers, digital billboards, kiosks, smart TVs, smartphones and glasses and contact lenses. As part of natural user interfaces it will help to make consumers' daily lives simpler, more efficient, safer, and more enjoyable. As interactive digital devices record what people look at, this will generate an enormous amount of information available about their visual behavior in their day-to-day lives. Then, big gaze data will be as common as Internet click-stream data has become in the past decades. That wealth of big gaze data will be of great value to market research companies, manufacturers, retailers and service providers, because it will enable them to better tailor user interfaces, products, services, prices, promotions and advertising to individual consumers' moment-to-moment interests. Dealing with that data, however, will require powerful cloud computing and novel analytical methods that are currently being developed. If consumer privacy is carefully considered, unprecedented new insights into consumers' visual behavior will become available, which hold the promise of massively improving managerial decision making and the consumer experience.

References

Aribarg, A., Pieters, R., & Wedel, M. (2010). Raising the BAR: Bias adjustment of recognition tests in advertising. *Journal of Marketing Research*, *47*(3), 387–400.

Ashby, N. J. S., Johnson, J., Krajbich, I., & Wedel, M. (2016). Applications and innovations of eye-movement research in judgment and decision making. *Journal of Behavioral Decision Making*, *29*(2–3), 96–102.

Atalay, A. S., Bodur, H. O., & Rasolofoarison, D. (2012). Shining in the center: Central gaze cascade effect on product choice. *Journal of Consumer Research*, *39*(4), 848–866.

Brasel, S. A., & Gips, J. (2008). Breaking through fast-forwarding: Brand information and visual attention. *Journal of Marketing*, *72*(6), 31–48.

Burton, L., Albert, W., & Flynn, M. (2014). A comparison of the performance of webcam vs. infrared eye tracking technology. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*, 1437–1441.

Buswell, G. T. (1935). *How people look at pictures: A study of the psychology and perception in art.* Chicago, IL: University of Chicago Press.

Chandon, P., Hutchinson, J. W., Bradlow, E. T., & Young, S. H. (2009). Does in-store marketing work? Effects of the number and position of shelf facings on brand attention and evaluation at the point of purchase. *Journal of Marketing*, *73*, 1–17.

Dodge, R. (1900). Visual perception during eye movements. *Psychological Review*, *7*, 454–465.

Dreze, X., & Hussherr, F.-X. (2003). Internet advertising: Is anybody watching? *Journal of Interactive Marketing*, *17*(4), 8–23.

D'Ydewalle, G., & Tamsin, F. (1993). On the visual processing and memory of incidental information: advertising panels in soccer games. In D. Brogan, A. Gale & K. Carr (Eds.), *Visual search 2* (pp. 401–408). London: Taylor & Francis.

Fitts, P. M., Jones, R. E., & Milton, J. L. (1950). Eye movements of aircraft pilots during instrument-landing approaches. *Aeronautical Engineering Review*, *9*(2), 24–29.

Hering, E. (1879). Der Raumsinn und die Bewegungen des Auges. In L. Herrmann (Ed.), *Handbuch der Physiologie der Sinnesorgane* (vol. 3, pp. 341–601). Leipzig: Vogel.

Huey, E. B. (1898). Preliminary experiments in the physiology and psychology of reading. *American Journal of Psychology*, *9*, 575–586.

Janiszewski, C. (1998). The influence of display characteristics on visual exploratory search behavior. *Journal of Consumer Research*, *25*(3), 290–301.

Javal, L. (1879). Essai sur la physiologie de la lecture. *Annales d'Oculistique*, *82*, 242–253.

Josephson, S. (2005). Eye tracking methodology and the Internet. In K. Smith, S. Moriarty, G. Barbatsis & K. Kenney (Eds.), *Handbook of visual communication, theory, methods and media*. New York, NY: Routledge.

Karslake, J. S. (1940). The Purdue eye-camera: A practical apparatus for studying the attention value of advertisements. *Journal of Applied Psychology*, *24*(4), 417–440.

Leven, W. (1991). *Blickverhalten von Konsumenten: Grundlagen, Messung und Anwendung in der Werbeforschung*. Heidelberg: Physica-Verlag.

Li, D., & Parkhurst, D. (2005, September). *Open-source software for real-time visible-spectrum eye tracking*. Paper presented at the Second Conference on Communication by Gaze Interaction—COGAIN 2006: Gazing into the Future. Turin, Italy.

Liechty, J., Pieters, R., & Wedel, M. (2003). Global and local covert visual attention: Evidence from a Bayesian hidden Markov model. *Psychometrika*, *68*(4), 519–541.

Nixon, H. K. (1924). Attention and interest in advertising, *Archives of Psychology*, *72*, 5–67.

Noton, D., & Stark, L. (1971). Scanpaths in eye movements during pattern perception, *Science*, *171*, 308–311.

Orschansky, J. (1899). Eine Methode die Augenbewegungen direct zu untersuchen. *Zentralblatt für Physiologie*, *12*, 785–785.

Pelz, J., Canosa, R., Kucharczyk, D., Babcock, J., Silver, A. & Konno, D. (2000). Portable eyetracking: A study of natural eye movements. In *Proceedings of the SPIE, Human Vision and Electronic Imaging* (pp. 566–582). San Jose, CA: SPIE.

Pieters, R., & Warlop, L. (1999). Visual attention during brand choice: The impact of time pressure and task motivation. *International Journal of Research in Marketing*, *16*(1), 1–16.

Pieters, R., Warlop, L., & Wedel, M. (2002). Breaking through the clutter: Benefits of advertisement originality and familiarity for brand attention and memory. *Management Science*, *48*(6), 765–781.

Pieters, R., & Wedel, M. (2004). Attention capture and transfer in advertising: Brand, pictorial and text size effects. *Journal of Marketing*, *68*(2), 36–50.

Pieters, R., & Wedel, M. (2007). Goal control of attention to advertising: The Yarbus implication. *Journal of Consumer Research*, *34*(2), 224–233.

Pieters, R., & Wedel, M. (2012). Ad Gist: Ad Communication in a Single Eye Fixation. *Marketing Science*, *31*(1), 59–73.

Pieters, R., Wedel, M., & Batra, R. (2010). The stopping power of advertising: Measures and effects of visual complexity. *Journal of Marketing*, *74*(5), 48–60.

Pieters, R., Wedel, M., & Zhang, J. (2007). Optimal feature advertising design under competitive clutter. *Management Science*, *53*(11), 1815–1828.

Poffenberger, A. T. (1925). *Psychology in advertising*. Chicago, IL: A. W. Shaw.

Radach, R., Lemmer, S., Vorstius, C., Heller, D., & Radach, K. (2003). Eye movements in the processing of print advertisements. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 9–42). Oxford: Elsevier Science.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422.

Rayner, K., Rotello, C. M., Stewart, A. J., Keir, J., & Duffy, S. A. (2001). Integrating text and pictorial information: Eye movements when looking at print advertisements. *Journal of Experimental Psychology: Applied*, *7*(3), 219–226.

Rosbergen, E., Pieters, R., & Wedel, M. (1997). Visual attention to advertising: A segment-level analysis. *Journal of Consumer Research*, *24*(4), 305–314.

Russo, J. E. (1978). Eye fixations can save the world: A critical evaluation and a comparison between eye fixations and other information processing methodologies. In H. K. Hunt (Ed.), *Advances in Consumer Research* (pp. 561–570). Ann Arbor, MI.: Association for Consumer Research.

Shi, S. W., Wedel, M., & Pieters, F. R. (2013). Information acquisition during online decision making: A model-based exploration using eye-tracking data. *Management Science*, *59*(5), 1009–1026.

Starch, D. (1923). *Principles of advertising*. Chicago, IL: A. W. Shaw.

Stewart, A. J., Pickering, M. J., & Sturt, P. (2004). Using eye movements during reading as an implicit measure of the acceptability of brand extensions. *Applied Cognitive Psychology*, *18*(6), 697–709.

Stüttgen, P., Boatwright, P., & Monroe, R. T. (2012). A satisficing choice model. *Marketing Science*, *31*(6), 878–899.

Teixeira, T. S., Wedel, M., & Pieters, R. (2010). Moment-to-moment optimal branding in TV commercials: Preventing avoidance by pulsing. *Marketing Science*, *29*(5), 783–804.

Van der Lans, R., Pieters, R., & Wedel, M. (2008a). Competitive brand salience. *Marketing Science*, *27*(5), 922–931.

Van der Lans, R., Pieters, R., & Wedel, M. (2008b). Eye-movement analysis of search effectiveness. *Journal of the American Statistical Association*, *103*(482), 452–461.

Van der Lans, R., Wedel, D., & Pieters, R. (2011). Defining eye-fixation sequences across individuals and tasks: The binocular-individual threshold (BIT) algorithm. *Behavior Research Methods*, *43*, 239–257. doi:10.3758/s13428-010-0031-2

Wade, N.J. (2010). Pioneers of eye movement research. *Perception*, *1*(2), 33–68.

Wedel, M., & Pieters, R. (2000). Eye fixations on advertisements and memory for brands: A model and findings. *Marketing science*, *19*(4), 297–312.

Wedel, M., & Pieters, R. (2008). Eye-Tracking for Visual Marketing. *Foundations and Trends® in Marketing*, *1*(4), 231–320.

Wedel, M., & Pieters, R. (2015). The buffer effect: The role of color when advertising exposures are brief and blurred. *Marketing Science*, *34*(1), 134–143.

Wedel, M., Pieters, R., & Liechty, J. (2008). Attention switching during scene perception: How goals influence the time course of eye movements across advertisements. *Journal of Experimental Psychology: Applied*, *14*(2), 129–138.

Yarbus, A. L. (1967). *Eye movements and vision*, New York, NY: Plenum Press.

Zhang, J., Wedel, M., & Pieters, R. (2009). Sales effects of attention to feature advertisements: A Bayesian mediation analysis. *Journal of Marketing Research*, *46*(5), 669–681.

42

The Quantified Self

Jeong-Ki Hong and Jun-Dong Cho

Scenario

After getting out of bed, the first thing I do in the morning is to check my sleeping time and habits through a smartphone application and see how well I've slept last night compared with the previous night. I then go for a morning run, wearing the smart band on my wrist, which can track the distance traveled, time taken, and calories burned during the exercise. The smart band application shows that I've already completed my current running plan, and suggests a new course with a longer distance. After the exercise, I come back home, take a shower, and stand on the scales. The scales inform me not only of my weight but also my body mass index (BMI) score, body fat percentage, and other information regarding my body. Furthermore, the scale gives me an analysis of the differences between the current and past data. While I prepare my breakfast, a screen on the refrigerator shows me my current nutritional health status. It is calculated based on last week's food intake and through a biosensor that senses my veins, collecting data regarding my cholesterol and blood glucose levels. After cooking the meal, I dip my spoon into a bowl, and suddenly the color of the spoon turns red, warning me that the content in the bowl could potentially exceed my daily sodium requirement. Maybe I've eaten more salt than usual. After finishing the meal, I go to my smart wardrobe to get dressed. Being aware of the weather conditions outside and my previous outfits, the smart wardrobe gives me suggestions on what clothes to wear. I also do not forget to wear my smart belt, which tracks my activity, mass, heart rate, HRV, and moods throughout the whole day. I get in the seat of my car. I know that I don't have to worry about being late to work since the navigation takes into account of the previous paths that I've taken and the current state of traffic to suggest a more suitable path. Before I start driving, I have time to check today's activity list, and know that I have cleared eight items out of 11 listed in my plan for today. Overall, in this week, I achieved in the average of 90% of the weekly plan and I wish to maintain the score in the following weeks.

The Wiley Handbook of Human Computer Interaction, Volume 2, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

The scenario above does not just show a future possibility: we can easily find most of these smart quantified-self tools, which are currently commercially available. They will become easier and more natural to use, so that we will not even recognize consciously the fact of being monitored by them.

The Concept of the Quantified Self

There is an old Chinese saying; "Know the enemy and know yourself; in a hundred battles you will never be in peril." Socrates also famously said, "Know yourself."

To know more, first we need to admit that we do not know anything. These two quotes encourage us to think about how much we know about ourselves. Some people might say they already know a lot about themselves; however, most of us do not have full access to the information about ourselves. In other words, until now, except for a few pioneers, we generally have not recognized the value of the information created by us.

Previously, there were two groups: those who made data and those who analyzed the data from the other group. However, nowadays these groups have been united, and thus those who create data can analyze it (Lee, 2013).

In particular, there are many information technology tools to conduct this analysis. We can thus quantify and measure almost everything that we can think of. For example, we may track and store the data generated from biological, physical, behavioral, environmental, and even psychological traits such as feelings and moods. We call this the quantified self (hereafter referred to as the QS), which involves quantifying information on ourselves or around us. The people who implement the QS, or try to find a way to achieve it, are referred to as "QSers," "Q-selfers," or "self-quantifiers."

Why do they want to quantify themselves?

Fundamentally, the QS is based on the belief that if we know more about ourselves in a measurable way, it will be easier to enhance our lives. Its short-term target is to quantify the information that is being generated from individuals, so that they can process the information into data that can be stored, shared, and analyzed as a progress map to which they can refer. Eventually they can obtain predictions and control of their bodies, as well as the motivation to improve their lives. Generally, motivations to engage in the QS are self-design, self-entertainment, self-association, self-discipline, and self-healing (Schreiber, 2013).

There are some disputes about viewing the QS as a scientific topic. However, no one would argue with the fact that the QS is creating an enormous ecosystem, including various technological areas that we shall discuss later in this chapter (Swan, 2013). The QS has been chosen as one of 10 rising technologies on the 2014 Global Economic Forum (Top 10 emerging technologies, 2014). Furthermore, the majority of the inventions and ventures tend to be related to this topic (Within the Internet of Things, 2013). There is no doubt that the QS is strongly related to various areas of HCI because the QS covers trendy HCI topics such as self-monitoring, wearable sensing, ubiquitous computing, big data, gamification, healthcare, and learning.

In this chapter we will explore the concept of the QS, starting with its history and the related areas of HCI, as well as examples in these areas.

Quantified Self-Movement

Gary Wolf and Kevin Kelly (Wolf, 2011b), the editors of *Wired Magazine*, introduced the QS movement. In 2007, they started a blog about the QS as a forum for people to meet and share their experiences of QS projects and develop their own QS technologies. Subsequently, many conferences and meetups, such as the "Quantified Self Show and Tell," began to emerge. This accelerated after Gary Wolf gave a speech on the QS at the 2010 TED (Wolf, 2010, 2011a). In fact, the QS movement is becoming significantly bigger every year (Mondry, 2014).

At TED, Wolf mentioned that new technologies, like sensors and devices, are changing our life patterns. He added that these tools should not only be used outside us but also inside us as a mirror and to control self-consciousness. In other words, the technology helps us to change our behavior in more efficient and productive ways (Wolf, Carmichael, & Kelly, 2010). This thinking led us to the concept of the "n = 1 experiment," which is an experiment on one's self. The advantages of this concept is for people to obtain the capability to care for themselves and to be able to conduct experiments using more novel and realistic approaches than the usual approaches of professional researchers. For specific examples, some of Bob Troia's experiments (Bob Troia, n.d.) have included glucose hacking and tracking, telomere analysis, bulletproof diet (cholesterol / bloodwork), and central nervous system (CNS) training.

Next, let us review how to develop a QS project.

Developing a QS Project

There are several ways of doing this; however, most QS projects or experiments have a certain framework with basic steps similar to those of a research experiment, according to Konstantin Augemberg of the blog Measured Me (Augemberg, 2012). First, we set the research hypothesis to verify, and we determine how to measure variables to verify the hypothesis. Next, we set the time to establish how long we will conduct the experiment to obtain enough data from ourselves. We should then choose a design to interpret the experiment's results, which would be the proper approach for the hypothesis. In the last step, we can test the data that we have collected by conducting a statistical test to evaluate whether the difference of the variables is significant. The data from the QS project might not be normally distributed or symmetric. The data usually provides small sample sizes from a single subject, so the data from the QS is not the same as other experimental data. Thus, he recommended more a robust, nonparametric test, such as a nonparametric version of Hedge's g test. (We omit the detailed explanation due to space limits. Please refer to Augemberg, 2012; Durlak, 2009; Olive & Smith, 2005; Schacht, Bogaerts, Bluhmki, & Lesaffre, 2008.)

The works we will introduce in this chapter might not perfectly fit into the common concept of the QS that we outlined above. However, in this chapter, we attempt to give a sense of emerging HCI and the QS by giving some examples of the essential QS components, such as quantifying human factors related to one's body, increasing motivation, recording data for a certain period to show progress, and trying to draw a conclusion, providing useful information from the quantified data.

Next, let us start by determining what kinds of information relates to the hypothesis we wish to quantify and track.

Life Logging: Self-Tracking

As mentioned previously, the QS is about tracking and recording information about ourselves and surrounding ours. This is why people usually use the term "QS" alternating with the term "life logging" or "self-tracking." Generally, the aim of the QS is as shown in Table 42.1.

As sensing technology is improving, the number of QS projects is increasing. That implies that there are more types of variables besides those shown in Table 42.1. Beyond information such as the number of walking steps or distance we traveled, we can also obtain information by tracking and storing data such as stress, moods, and specific behaviors in real time through various analysis methods.

Speaking of stress and mood, a recent review from Kanjo and Chamberlain (2015) discusses pervasive affective sensing. Pervasive affective sensing is about tracking users' feelings in a pervasive way by tracking their affective states using various tools, including emotion-monitoring systems, analysis methods, and applications.

There are various measurements besides self-reporting to capture our emotions, such as physiological signals, facial expression, speech, phone usage, social networks, and mobile network data. Self-reporting is a relatively more manual way of tracking, while other measurements can be collected automatically with various tools. The analysis method depends on what kind of the measurement the researcher is using such as physiological signal-based analysis or text-based analysis.

By monitoring and analyzing our feelings, we can also use the data to develop various applications, which are used to track mental health. For example, Empath (Dickerson, Gorlin, & Stankovic, 2011) is a monitoring system that detects and tracks depression symptoms in real time. The system consists of wireless sensors, a touch screen, and a mobile device to capture users' behavioral data, such as speech, sleep, weight, and movement data. The application of affective sensing can also be helpful for

Table 42.1 Quantified self-tracking categories and variables.

Category	*Variables*
Consumption	Calories, caffeine, sugar, water
Psychological state	Stress, mood, tension, relaxation, hunger
Physiological state	Heart rate, cholesterol, fat, weight, blood pressure, body temperature, body PH
Physical activity	Sleep, running, walking, swimming, sitting, watching something
Environment	Weather, location, place

urban planning. "Sensing the city" (Bergner, Exner, Zeile, & Rumberg, 2012) shows the potential of using physiological and psychological data to design urban spaces.

As the number of methods to track variables is increasing, the methods are becoming more sophisticated. For instance, McDuff, Gontarek, and Picard (2014) show remote changes of physiological parameters using an automated prediction system consisting of a digital camera and software to measure cognitive stress.

Teng Fu and Allan MacLeod (2014) also proposed a robust and low-cost FSR sensor-based system that can recognize the sitting posture, activities, and behaviors of the user on the chair.

Now, including the methods mentioned above, what kind of tools can we use for the QS in general?

Internet of Things (IoT)

Even though not every QS method relies on electronic devices or the Internet, as the number of technologies related to the IoT increases, the number of QS methods is also rapidly increasing. There are several definitions of IoT, but according to the U.S. National Intelligence Council, "The 'Internet of Things' is the general idea of 'things', especially everyday objects that are readable, recognizable, locatable, addressable, and controllable via the Internet" (Marien, 2013). Internet-accessible devices, such as laptops, desktops, smartphones or tablets that are connected to the Internet are especially suitable for the QS project. Through the IoT we can track and record our status continuously over 24 hours even when we are not aware of its presence and we can also easily turn existing devices into QS tools. For example, an IoT weighing machine connected to the Internet can not only weigh but can also upload the data onto a cloud and analyze it automatically without extra effort.

Smartphone

As sensors such as accelerometers, gyro sensors, or magnet sensors inside phones have become smarter and smaller, the smartphone has become one of the easiest tools to use for the QS (Barcena, Wueest, & Lau, 2014).

The work of Jariyasunant et al. (2015) is an example of using a smartphone as a tool to quantify the users' data and change their behaviors. Jariyasunant et al. designed and evaluated a system named quantified traveler (QT), which is a computational travel feedback system. Travel feedback is an established programmatic method with which travelers record their travels in diaries and meet with a counselor to obtain advice about their trip. However, counselor costs and efforts of recording travel data by travelers are problematic. Hence, this system tracks travel data in a relatively automated way via a smartphone application.

This project fits the QS quite well in terms of quantifying the data of our activity through a smartphone. Moreover, there is a feedback system, which is website in this case, by analyzing and sharing the data to motivate behavior changes of people.

Beyond this project, there are many self-tracking applications, and the number of applications is increasing. We list some of the applications In Table 42.2.

Table 42.2 Some self-tracking applications.

Name	*Description*	*Reference site*
80 Bites	Tracking how much food you ate during a day.	https://play.google.com/store/apps/details?id=com.eightybites
Moodscope	Measuring your mood everyday with a card game and track your daily scores to learn what could be causing your ups and downs.	https://www.moodscope.com
Mind Bloom	Simple social game aimed at improving your quality of life by rewarding you for certain behavior.	http://www.mindbloom.com
Waterlogged	Recording your water intake during a day.	https://itunes.apple.com/us/app/waterlogged-drink-more-water/id352199775?mt=8
Runkeeper	An application that track the time, distance, speed, pace, calories, and path on a map of your fitness activities.	https://itunes.apple.com/us/app/runkeeper/id300235330?mt=8
Facebook Timeline	SNS service that collects enormous amount of your activity that can be analyzed.	https://www.facebook.com

Wearable Device

These days, wearable sensors and devices are more popular QS tools than smartphones. This is not only because many wearable devices are emerging but also because wearable devices are quite suitable for wearing on our body to track data (Appelboom et al., 2014).

One example is Yongwon Jang (Jang et al., 2011) who developed a waist belt-type smart device that can track calories and waist circumference changes.

The main purpose of this belt is to serve as integrated hardware, which includes an accurate and automatic calorie tracker and a waist-measuring module. Long-term monitoring of data gathered from the device can also give a sense of the relationship between obesity and daily life patterns to predict and prevent obesity.

Now, lets turn to HCI topics related to the QS.

Big Data

Big Data and the QS are closely related to each other. In fact, the enormous data that the QS generates are fascinating for most data analysts. As Wolf mentioned, if the first stage of the QS was a personal experiment that individuals could manage by themselves, the QS is now expanding through the Internet and cloud. Moreover, as the number and the time of tracking variables are growing fast, the data that one person generates is in a geometrical progression. For example, if a heart rate monitor tracks

250 samples in 1 s, this means the monitor will generate data of over 9 gigabytes per person in a month. If there are more variables that the person is tracking, the size of the data will be easily doubled or trebled (Swan, 2013). When these data from each person are uploaded on the Internet, the data then truly becomes Big Data.

This relationship gives certain advantages not only for analysts or marketers but also for the individual who is conducting the QS.

First, individuals gain personalized and advanced information for themselves, which the analyst recreated from the raw data (Barcena, Wueest, & Lau, 2014). We may use one of the devices we previously mentioned to upload the data generated by the device and we let other professionals use those data to provide various information, such as the distance or the time you ran, when you usually run, or where you run or live. If you upload your demographic information on the application, which matches with the device, this data becomes quite useful to market analysts. By using these data, they can offer you personalized information, such as certain exercises for working out, running courses, or well-timed advertisements, such as new shoes you might be interested in or sales information based on your current location.

In addition, on major health websites, individuals upload their experiences, symptoms, and treatment on the Internet so that they can gain more statistical and subjective information (Swan, 2009).

By sharing personal data and results, individuals can have the opportunity to compare their information with others and make their QS more developed, although the data might not be Big Data.

Gamification

Gamification is one of the key words in the QS because the data generated from the QS effectively creates a game (Whitson, 2013). The gamification increases the overall motivation by using a game mechanism, such as by rewarding users with PBL (point, badge, and level) according to the progress that person has achieved (Deterding, Dixon, Khaled, & Nacke, 2011). To apply gamification, it is important to digitize actions (Marczewski, 2013), which indicates the close relationship of gamification and the QS. Furthermore, since the game mechanism can easily modify feedback to change the direction of a person (Ellerbrok, 2011), there is a certain advantage to using gamification for the QS.

The concept of "play" in gamification can also help people to alleviate boredom by recording themselves (Ellerbrok, 2011). One of the most successful examples of combining the QS and gamification is Nike + .

Nike + did not start by simply tracking distance via GPS and showing data through statistics and graphs but started to design a whole system in a more playful way. You can view this through the marketing video of Nike+. By adding game sounds, such as inserting coins, and rewarding users, such as with badges and trophies, according to achievements, it increases flow and motivation. With the records uploaded on the Internet, the user now does not run alone, but with many other people, creating a sense of community simultaneously. These game factors make the overall experience of running more enjoyable and compensate for the physical pain of running eventually motivating people to run again with the device.

Of course, not every QS should adopt gamification, and not every QS is able to adopt it. Gamification should be designed to enhance motivation in a natural way; otherwise, it can be counterproductive (Webb, 2013).

Let us move onto some applications.

Application

So far, we have explored various topics and examples related to the QS with the approach of HCI. In this section we will introduce some concept design ideas that have low fidelity but are more novel and fresh.

We will introduce some research projects from a recently established graduate program, named "human ICT (information and cognitive technology) convergence" at Sungkyunkwan University in Korea. As we can see from the words "human," "ICT," and "convergence" in the department name, the students in the department have various backgrounds such as computer science, design, engineering, psychology, history, or literature. Multidisciplinary convergence therefore tends to go beyond knowledge-based limitation.

In the following sections we will show several projects performed from the department's H-Lab. (i.e., humaneering laboratory), which we might regard as examples of the QS.

Smart wig

Through the smart wig, Kim et al. (2014) tried to develop a wearable health-monitoring device that can track certain circumstances and events that happen to a patient, and especially one who has cancer.

The device has three main functions. First, the smart wig monitors patients' specific physical states and is particularly focused on fall detection. Secondly, it records physiological data, such as heart rate and body temperature, and finally situational information such as current location. To implement these functions, the prototype consists of three parts: (a) the input module, which collects data from patients, (b) the microcontroller unit, which processes and communicates with a smartphone application, and (c) the output module, which rings an alarm. The input module includes an accelerometer, temperature sensor, and pulse sensor, and the output module includes an LED and buzzer. The algorithm is coded onto the MCU part. In particular, the fall detection solution of the project can identify four different stages related to falls: normal, dynamic transition, analogous falling, and falling.

The relevant smartphone application allows the supervisor, such as a doctor or a nurse, to monitor the patient's state and provides an alarm for emergencies. The application obtains information from the smart wig wirelessly and displays the physiological data.

The project above only contains a function for capturing irregular moments, whereas the next project focuses on how to track and express information gathered about our bodies during the tracking period.

Flower-shaped avatar

In the flower-shaped avatar project for better posture awareness (Hong, Song, Cho, & Bianchi, 2015), we tried to develop a platform to track the user's back posture and the circumstances in which the user worked at a desk by using various sensors, including a wearable device. We use an ambient display and a gamification application to give feedback to the user.

We started this project because back-related health problems have become a serious issue. People usually spend a lot of time working in the sitting position, so incorrect postures and long sitting sessions are among the main causes of back pain and discomfort (Hakala, Rimpelä, Saarni, & Salminen, 2006). Hence, one of the key objectives of this project is to quantify the user's posture and additional information, such as closeness to a monitor or sitting time.

The back-posture tracking device includes a gyro sensor that can be attached to the user's back to track how much the user bends it. We divided the user's back posture into five different states according to bend level, and gathered extra information on how many times the user changes posture or how long the user has been on the same spot. The second purpose of this work is to visualize the quantified information tracked with the sensors, and adopt a flower-shaped avatar for the feedback system.

The flower-shaped avatar consists of a set of 3D-printed flower, bendable stem, and flower pot. Inside the pot, there is a motor, a microcontroller, and a speaker.

The feedback mechanism works in the following way. If users bend their backs, the flower-shaped avatar bends its stem to mimic the user's back posture, indicating that the user's back posture is in incorrect position.

If the user has been sitting for a long time without intermission, the color of the stem will turn blue to indicate that the user needs to stretch or take a walk.

In addition, a simple gamified application with a digital flower avatar is developed for smart phone to communicate with the wearable device that influences the state of the user's back posture (Hong, Koo, Ban, Cho, & Bianchi, 2015). The user can therefore nurture both physical and digital avatars by sustaining a good posture.

This project is an example of the QS using an ambient display and gamification.

Gait-tracking device

Another QS-related project is the gait-tracking device implemented by Shin, Lee, Kim, Bae, and Cho (2014). There are many studies saying that inappropriate gait might unbalance the pelvis, spine, and joints, and if it is maintained, serious illness can occur (Go, Hong, Lee, & An, 2013). Although people are aware of the important effects of gait balance on health, most people do not care about it because equipment that measures gait is quite expensive and is physically and psychologically demanding (Xu et al., 2012).

To overcome these issues, this project designs a feasible prototype of a wearable device that tracks the gait of the user, and suggests an application design that uses statistical analysis and various feedbacks to increase the motivation of the user to change walking behavior.

The footwear prototype consists of the Arduino Mini, a gyro and accelerometer sensor, an FSR sensor, and a Bluetooth module. The prototype particularly tracks abnormal gait such as toe-in walking and toe-out walking. The algorithm for discriminating the gait is based on the work of Rosenbaum (2013). It used five FSR sensors to analyze the pressure distribution under the feet and used an accelerometer and a gyro sensor to measure the distance between the feet. It then merged those data to judge the gait of a user and sent the data to a smartphone application server through a Bluetooth connection.

The novelty of this work is that the application is designed so that all essential information is provided in fewer frames, to appear more intuitive to a user. The application has three screens. The first screen shows the amount of walking and period of each gait using a circular diagram and coloring. The colored outer line of the circle indicates the gait types, and the graph inside the diagram shows the amount of exercise. The second screen also shows a statistical bar graph of one week with a goal bar to motivate the user. On the third screen, there is a graphical gamified motion of feet walking on a lawn that shows that the more the user tends to adopt a bad gait, the feet on the screen tend to walk over the flowers, which implies improper gait.

Those projects so far follow the basic format of the QS project with IT technology, which has tracking with a wearable device, analysis of statistics, and motivation through visualization and gamified feedback. The next projects exhibit more focused visual data from our body.

Lily Kickee

The project "Lily Kickee" developed by Been, Sanghoo, Jaewon, Byung-chull, and Jun-dong (2015) involves a wearable device for pregnant women. The device shows fetal movements with an intuitive and graphical animation so that pregnant women and related people around them can feel close to a fetus. In this case, fetal movement corresponds to the data from their body.

Fetal movement is an important indicator of the health of the fetus. A decrease in fetal movements might result in a dangerous situation regarding the health of the fetus and the pregnant woman. Thus, recording fetal movement is one way of examining the fetus and can reduce the possibility of stillbirth significantly (Winje, Røislien, & Frøen, 2012).

The Lily Kickee prototype uses Lilypad Arduino. Buechley and Eisenberg (2008) created this smart wear. It is small, light, and easy to connect to other sensors through the conductivity thread. Lily Kickee has FSR sensors to detect fetal movement, an LED strip to express the movement, and a 3.7-V Lipo rechargeable battery. The three FSR sensors are located on the left, center, and right of the user's belly where the fetal movement usually occurs and those LEDs are designed as the shape of a baby footprint. They are now pushing this work forward to examine the effects and the advantages of using the prototype continuously for a certain period, which makes the project more related to the QS.

If you consider these projects, you can easily find that most of the works is to implement small sensors, such as an accelerometer, gyro sensor, and FSR sensor through IoT connections, which we explained previously. More importantly, novice graduate students with various undergraduate backgrounds (such as IT, graphic or product design, engineering) are able to perform this work, implying that developing the QS with an IT (information technology) approach has become much easier than in the past.

Spoon up

The last project we want to introduce is more conceptual and is designed for future QS. The name of the project is "Spoon Up" by Noh, Shin, Cho, and Bianchi (2014), and it is about a smart spoon that can monitor the properties of the food that a baby consumes.

In this work, Noh et al. propose a conceptual prototype design based on guidelines derived from the results of conducting a formative study, including an in-depth interview and speed dating to form the guidelines and design.

Considering the user's needs, Noh et al. proposed a conceptual design of a device that has a rubber head and an infrared sensor, which allows users to analyze the nutrients on the spoon. A temperature warning is displayed using an LED on the spoon handle. The developed smartphone application's main interface is divided into three sections: temperature, nutrition visualization, and a diary.

This work illustrates that when we design QS technology we need to design it carefully by considering and understanding the user's needs.

The visual aids of some of projects described above and also other projects that have been developed from our lab are available on YouTube (https://www.youtube.com/channel/UCRhvKxj-aQoBDxTodJzmP3w/videos?shelf_id=0&view=0&sort=dd).

As previously mentioned, QS technology is becoming relatively easy and convenient to develop, but has disadvantages and limitations, which we shall elaborate in the following section.

Limitations

As the size of QS technology is increasing, the limitations of QS are receiving attention.

The biggest risk and limitation of QS technology is the security of the information. This risk has grown rapidly because with the IoT we can easily upload information through the Internet. According to a survey conducted by Pew Research and Carnegie Mellon University, published in September 2013 (Rainie et al., 2013), in the United States, over 86% of adults tried to protect their privacy when they were on the Internet. This statistic shows that people are actually concerned about their privacy. Education movements that help people to use tools to protect their privacy, such as the Crypto Party movement (Park, 2012) share this concern.

By failing to protect personal data, several issues can occur. By using your demographic information, some people might commit crimes, such as identity theft (Barcena, Wueest, & Lau, 2014).

Apart from issues of privacy, there is also the possibility of confusion due to the absence of a standard quantifying system. For example, although there are many smart wristbands that track your walking steps, each band uses a different algorithm in its device, or some of them track inaccurately (Linning, 2015) so the data from devices cannot be compared in a unidimensional way. However, some websites and software have been developed to solve this problem, to compute the data, and draw a standard analysis.

The last concern about QS technology is that the QS might not be a solution for enhancing life, but a distraction that leads to narcissism (Cohen, 2014; Teitell, 2012). This is why we should build QS projects with clear targets and hypotheses, to avoid losing sight of the goal of conducting QS and ending up with numbers that do not mean anything.

Conclusion

We might see the QS as a trend rather than a term that has a fixed definition. In the near future, it will become involved in our lives more deeply, as in the scenario at the beginning of this chapter. People who are interested in exploring the QS currently include early adopters, fitness, technology, and personal development enthusiasts, biohackers, and patients suffering from various health problems. In the future it will extend to a much wider group of people. Moreover, the topic of the QS will not only be limited to healthcare or self-development, but also to new areas such as learning.

Eventually, QS technology will become very useful, but in a natural manner so that we do not notice it, measuring human factors, and enhancing our unconscious awareness of our bodies and environment with self-knowledge that will affect our lives in various ways.

References

Appelboom, G., Yang, A. H., Christophe, B. R., Bruce, E. M., Slomian, J., Bruy, B. O.,… Connolly, E. S. (2014). The promise of wearable activity sensors to define patient recovery. *Journal of Clinical Neuroscience*, *21*(7), 1089–1093.

Augemberg, K. (Nov. 2012). Quantified self how-to: Designing self-experiments: Measured me. *Humanity Plus*, retrieved from http://hplusmagazine.com/2012/11/14/quantified-self-how-to-designing-self-experiments/

Barcena, M. B., Wueest, C., & Lau, H. (2014). *How safe is your quantified self?* Symantech: Mountain View, CA.

Bergner B. S., Exner J.-P., Zeile, P., & Rumberg, M. (2012). Sensing the city—how to identify recreational benefits of urban green areas with the help of sensor technology. In M. Schrenk, V. Popovich, P. Zeile, & P. Elisei (Eds.), *Proceedings of REAL CORP 2012: Remixing the City* (pp. 737–746). Vienna, Austria: CORP.

Buechley, L., & Eisenberg, M. (2008). The LilyPad Arduino: Toward wearable engineering for everyone. *IEEE Pervasive Computing*, *7*(2), 12–15.

Cohen, J. (2014, February 5). Quantified self: The algorithm of life. *Prospect Magazine*. Retrieved from http://www.prospectmagazine.co.uk/arts-and-books/quantified-self-the-algorithm-of-life

Deterding, S., Dixon, D., Khaled, R., & Nacke, L. (2011, September). From game design elements to gamefulness: defining gamification. In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments* (pp. 9–15). New York, NY: ACM.

Dickerson, R. F., Gorlin, E. I., & Stankovic, J. A. (2011, October). Empath: a continuous remote emotional health monitoring system for depressive illness. In *Proceedings of the Second Conference on Wireless Health* (p. 5). New York, NY: ACM.

Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. *Journal of Pediatric Psychology*, *34*(9), 917–928.

Ellerbrok, A. (2011). Playful biometrics: controversial technology through the lens of play. *The Sociological Quarterly*, *52*(4), 528–547.

Fu, T., & Macleod, A. (2014). IntelliChair: An approach for activity detection and prediction via posture analysis. In *International Conference on Intelligent Environments (IE)* (pp. 211–213). Shanghai: IEEE.

Go, E., Hong, S., Lee, K., & An, K. (2013). Effect of active change of foot progression angle on lower extremity joint during gait. *Korean Journal of Sport Biomechanics*, *23*(1), 85–90.

Hakala, P. T., Rimpelä, A. H., Saarni, L. A., & Salminen, J. J. (2006). Frequent computer-related activities increase the risk of neck–shoulder and low back pain in adolescents. *European Journal of Public Health*, *16*(5), 536–541.

Hong, J. K., Koo, B. C., Ban, S. R., Cho, J. D., & Bianchi, A. (2015, September). BeuPo: A digital plant that you can raise and customize with your current posture. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers* (pp. 1015–1020). New York: ACM.

Hong, J. K., Song, S., Cho, J., & Bianchi, A. (2015, January). Better posture awareness through flower-shaped ambient avatar. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction* (pp. 337–340). New York: ACM.

Jang, Y., Noh, H. W., Lee, I. B., Song, Y., Jang, W. I., & Lee, S. (2011). Development of an integrated obesity management waist belt system composed of calorie tracking and waist circumference measuring module for long term monitoring. In *Engineering in Medicine and Biology Society (EMBC, 2011), Annual International Conference of the IEEE* (pp. 2172–2175). Boston, MA: IEEE.

Jariyasunant, J., Abou-Zeid, M., Carrel, A., Ekambaram, V., Gaker, D., Sengupta, R., & Walker, J. L. (2015). Quantified traveler: Travel feedback meets the cloud to change behavior. *Journal of Intelligent Transportation Systems*, *19*(2), 109–124.

Jeon, B., Park, S., Cho, J., Bae, B., & Cho, J. (2015, Sep.). Smart Maternity Clothes for visualizing fetal movement data. In *Adjunct Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers* (pp. 189–192). Osaka, Japan: ACM.

Kanjo, E., & Chamberlain, A. (2015). Emotions in context: examining pervasive affective sensing systems, applications, and analyses. *Personal and Ubiquitous Computing*, *19*(2), 109–124.

Kim, N., Ryu, J. H., Park, M., Keum, J., Yoo, J., Lee, J., & Cho, J. D. (2014, December). Smart wig for patient with chemotherapy induced alopecia: Emphasis on fall detection and emergency alert. In *Proceedings of HCI Korea* (pp. 246–252). Seoul, Korea: Hanbit Media, Inc.

Lee, V. R. (2013). The quantified self (QS) movement and some emerging opportunities for the educational technology field. *Educational Technology*, (November–December), 39–42.

Linning, S. (2015, March 4). Why your fitness tracker could be a waste of money: New research shows the must-have gadgets are often wildly inaccurate by as much as 40 per cent. *Daily Mail*. Retrieved from http://www.dailymail.co.uk/sciencetech/article-2914523/Why-fitness-tracker-waste-money-New-research-shows-gadgets-wildly-inaccurate-40-CENT.html

Marczewski, A. (2013, March 25). Feedback loops, gamification and employee motivation [Web log post]. Retrieved June 1, 2015, from http://www.gamified.uk/2013/03/25/feedback-loops-gamification-and-employee-motivation/

Marien, M. (2013). *Global trends 2030: Alternative worlds*. Washington, DC: National Intelligence Council.

McDuff, D., Gontarek, S., & Picard, R. (2014, August). Remote measurement of cognitive stress via heart rate variability. In *Proceedings of the 36th Annual International Conference Engineering in Medicine and Biology Society (EMBC)* (pp. 2957–2960). Chicago, IL: IEEE.

Mondry, B. (2014, August 25). What does the "quantified self" movement mean for health care researchers? *Quirk's Media*. Retrieved from http://www.quirks.com/articles/2014/20140825-1.aspx?searchID=1074409942&sort=5&pg=1

Noh, G., Shin, J., Cho, J., & Bianchi, A. (2014). Guidelines for a baby's nutrition monitoring system. In *Proceedings of HCI Korea* (pp. 92–97). Seoul, Republic of Korea: Hanbit Media, Inc.

Olive, M. L., & Smith, B. W. (2005). Effect size calculations and single subject designs. *Educational Psychology*, *25*(2–3), 313–324.

Park, A. (2012, October 16). Cryptoparties teach data privacy to the public. *SBS*. Retrieved from http://www.sbs.com.au/news/article/2012/10/16/cryptoparties-teach-data-privacy-public

Rainie, L., Kiesler, S., Kang, R., Madden, M., Duggan, M., Brown, S., & Dabbish, L. (2013). *Anonymity, privacy, and security online*. Retrieved from Pew Research Center website: http://pewinternet.org/Reports/2013/Anonymity-online.aspx

Rosenbaum, D. (2013). Foot loading patterns can be changed by deliberately walking with in-toeing or out-toeing gait modifications. *Gait and Posture*, *38*(4), 1067–1069.

Schacht, A., Bogaerts, K., Bluhmki, E., & Lesaffre, E. (2008). A new nonparametric approach for baseline covariate adjustment for two-group comparative studies. *Biometrics*, *64*(4), 1110–1116.

Schreiber, A. (2013, December). *Quantified self: Self tracking for health*. Paper presented at the Fourth International TEMOS Conference, Bonn.

Shin, J., Lee, H., Kim, S., Bae, B., & Cho, J. (2014). Visual feedback system design for gait management. In *Proceedings of HCI Korea* (pp. 286–291). Seoul, Republic of Korea: Hanbit Media, Inc.

Swan, M. (2009). Emerging patient-driven health care models: an examination of health social networks, consumer personalized medicine and quantified self-tracking. *International Journal of Environmental Research and Public Health*, *6*(2), 492–525.

Swan, M. (2013). The quantified self: fundamental disruption in big data science and biological discovery. *Big Data*, *1*(2), 85–99.

Teitell, B. (2012, May 8). The "quantified self" movement grows, but what are we really learning about ourselves? They've got your numbers. *Boston Globe*. Retrieved from http://www.bostonglobe.com/lifestyle/2012/05/07/quantified-self-movement-grows-but-what-are-really-learning-about-ourselves-they-got-yournumbers/VGDRdTtsdUyOAv6FvoYM2H/story.html

Top 10 emerging technologies for 2014. (2014, September 1). *World Economic Forum*. Retrieved from https://agenda.weforum.org/2014/09/top-ten-emerging-technologies-2014/

Troia, B. (n.d.). *N=1 experiments (bulletproof diet, intermittent fasting, blood glucose regulation and more), episode 22, the quantified body* [Podcast]. Retrieved from https://thequantifiedbody.net/n1-experiments-bob-troia/

Webb, E. N. (2013). Gamification: When It works, when it doesn't. In *Proceedings of Second International Conference, DUXU 2013, Held as Part of HCI International 2013* (pp. 608–614). Berlin: Springer.

Whitson, J. R. (2013). Gaming the quantified self. *Surveillance and Society*, *11*(1/2), 163–176.

Winje, B. A., Røislien, J., & Frøen, J. F. (2012). Temporal patterns in count-to-ten fetal movement charts and their associations with pregnancy characteristics: a prospective cohort study. *BMC Pregnancy and Childbirth*, *12*(1), 124.

Within the Internet of Things, Quantified Self Startups Grab $318M from VCs in the Last Year. (2013, November 19). *CB Insights* [Web log post]. Retrieved September 19, 2017, from https://www.cbinsights.com/blog/quantified-self-venture-capital/

Wolf, G. (2011a). Gary Wolf: Quantify yourself [Web log post]. Retrieved from http://tedx.amsterdam/2011/11/gary-wolf-quantify-yourself/

Wolf, G. (2011b, March 3). What is the quantified self [Web log post]. Retrieved September 28, 2017, from http://quantifiedself.com/2011/03/what-is-the-quantified-self/

Wolf, G., Carmichael, A., & Kelly, K. (2010). The quantified self [Video file]. Retrieved from http://www.ted.com/talks/gary_wolf_the_quantified_self.html

Xu, W., Huang, M. C., Amini, N., Liu, J. J., He, L., & Sarrafzadeh, M. (2012). Smart insole: A wearable system for gait analysis. In *Proceedings of the Fifth International Conference on Pervasive Technologies Related to Assistive Environments* (p. 18). New York, NY: ACM.

43

An HCI Approach in Contemporary Healthcare and (Re)habilitation

Anthony Brooks

Introduction: Apophasis / Preamble / Caveat

The research in this chapter originated in Wales. It arrived in Denmark in 1992 following 6 years in the United States. Denmark was selected because of its welfare system, which supports disabled people. The research has also been based in southern Sweden (Lund) for over 2 years, which has a similar Scandinavian perspective on the welfare of its people. The CAREHERE European project originated from the author's research.

Stimulating a person to experience such positive human traits as creating, playing, competing, enjoying, and communicating, is at the core of the concept. Such experiences are targeted in digitally enhanced environments where the stimuli can be selected and controlled to match the person's profile and the professional health worker's aims for that person's progress. Empowering people to manipulate the stimulus is the catalyst for interactions that engage them at a profound level. Participant interactions change from conscious to unconscious so that the intentional transforms into the unintentional, thus, at the desired state, the interactions are not consciously intended but rather they are automatic bodily responses to the stimuli manipulation immediately preceding a moment in time. Such interactions and the participant's responses inform the facilitator—the person controlling a session in a treatment program. Both qualitative and quantitative data are obtained from the interactions, which inform the design of subsequent digitally enhanced environments and tailoring and fine tuning to maintain participant progress.

The author's research activities have been philanthropic and self-funded under a not-for-profit / reinvestment strategy where a focus has consistently been on the research and development of apparatus and method to improve the life of people with impairment. The strategy targets those most marginalized so that system adaption (tailoring and tuning to an individual's profile) can address other participants with higher function in treatment programs. Next-generation health professional education is also targeted. Included in this reasoning is that comfort with digital

The Wiley Handbook of Human Computer Interaction, Volume 2, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

media is required to optimize "in-action" and "on-action" reflection and change to give effect to the session experiences (of both participant and facilitator) as well as the design and redesign iterations of the media content that is interacted with (e.g. virtual reality, games, auditory feedback, robotic lighting—as outlined in this chapter).

Included in future work is the use of digitization to impact testing apparatus and methods, and to enable mixed methods data outcomes to support use evaluations.

"SoundScapes" was the title given to the work in 1994. The original title of "Handi-MIDI" was not regarded as politically correct by the author, who conceived all aspects of the work and whose concept originated the research. "Personics" where mentioned in this chapter is a product appropriated by the Personics Company, Aarhus, Denmark, from the author's SoundScapes research, which forms the basis of the patent family assigned to Personics where the company's CEO is listed as "coinventor." The Personics product and company no longer exist and are not associated with this author.

History and Background

Family disability leading to tacit knowledge

This chapter focuses on a body of mature work that evolved from the author's childhood in close contact with family members who had profound impairments. One of the family members, the author's uncle, suffered from severe cerebral palsy and could barely move or verbalize. He could not feed, bathe, or look after himself in any way. He breathed, and his head and eyes moved with effort. Beyond this, his range of motion was minimal side-to-side movement of the upper torso, limited by severe spasticity. (Spasticity relating to altered skeletal muscle performance with paralysis, increased tendon reflex activity and hypertonia, which is typically caused by an imbalance of signals from the central nervous system—brain and spinal cord—to the muscles. The condition is represented by an involuntary tightening of the person's torso / limbs, and thus is linked to velocity-dependent muscle tone, which results in a resistance to movement.) It was found that by positioning the uncle's wheelchair elbow rests in a precise position, a side-to-side motion could be initiated (partly "controlled") by the uncle alone. Importantly, his intentional downward pressure on each resting post allowed a greater range of motion without any help from others.

Inspired by his uncle's love of music, the author set out to find a means for alternative control of feedback via residual functional. Quality of existence was sought through the uncle being empowered to do something himself that was meaningful to him.

Concept creation, development and early empirical research Effect pedals allow creative sonic self-expression where tone variations are typically controlled by the travel of a moveable 5 V transport rocker assembly. Effects include volume, fuzzbox / distortion, envelope-follower, octave generator, wah-wah, overdrive, feedback / sustain, distortion, fuzz, EQ, and tremolo. Mixing and matching the signal chain order of multiple pedals offers almost infinite options in sound coloring; however, optimization is only achievable with experienced planning of the signal chain with knowledge, skill and competence in linking, aligned to extreme care because of potential sound degradation and consideration of safety (i.e. ensuring high-quality devices with robust electronics and housing). In other words, these electronic devices traditionally enable a performer to alter media (e.g. a musical instrument output, or other audio source, including vocal via microphone, etc.); however, quality of build can affect optimized performance and enhanced user experiences of any interactions.

In the case of the author's uncle, with pedals attached to the wheelchair's armrests, residual functional movement controlled musical amplitude and tone, resulting in observed empowerment and a fun experience.

Human gestural control—initial phase Effect control depends on human gestures. For pedals, typically a foot is used for real-time control so that hands are free to perform (e.g. guitar, keyboard, etc.); other effect devices are rack mounted with associated foot pedal mapping to allow remote control. Device "programming" determines the sound coloring being controlled and is based upon parameter change adjustment—often knobs on hardware pedals and rack-mounted devices—which are then more finely adjusted by real-time gesture adjustment. Contemporary music software is commonly interface to a digital audio workstation (DAW) or even tablets, where often graphical virtual sliders, knobs, and other controllers replicate hardware devices. However, even in recent systems development, it is common that control via human gesture is preferred when tactile response is evident i.e. the feed-forward input giving tactile / haptic feedback matching the manipulated content feedback (i.e. audio). Some effects subtly "color" a sound whereas others transform it dramatically. Experiments with disabled volunteers illustrated potential beyond foot / torso as a tool to supplement therapy through an alternative approach to empower control (Brooks, 1999).

Human Gestural Control—Evolved Phase During the early phases of the research it was clear that whilst the manipulation causality showed potential, access to the devices was limited—they had to be adapted to each individual preferences for access, for example with an attachment to wheelchair armrests (which often became dislodged). Alongside mechanisms to improve access, there was a clear need to evaluate their potential through evidence of benefit from the interactions beyond subjective observations.

The research subsequently investigated invisible sensor interfaces, which eventually realized a bespoke infrared gesture-sensing system that was based upon attaching individual sensors on the end of a flexible gooseneck light holder. This assembly enabled one to be able to source both minuscule motions (even down to movements that were hard to detect with the eye) and gross motions (e.g. limb, torso, full-body motions). Whilst these offered no tactile feedback to the participant in order to mirror / associate with the manipulated feedback stimulus, the motion was mirrored by the content in such a way that a conduit between data source and stimulus received became attainable from the body's natural sense of proprioception and kinetic motion. This causal loop is referred to as *human afferent-efferent neural feedback loop closure* (Brooks, 2011; Brooks, Hasselblad, Camurri, & Canagarajah, 2002). This can be regarded as a "closed-loop" system of sensation, "decision," and reactions, where processing is carried out through the activity of afferent neurons (sensory neurons), interneurons, and efferent neurons (motor neurons). Simplified, human-received (sensed) stimuli is sourced, perceived and processed by the input conduits of the brain's afferent neural mechanisms to evoke the central nervous system reaction via efferent neural conduits, to initiate motor action via signal transmissions from the nerve cell body that carries nerve impulses away from the central nervous system toward the peripheral effector organs (mainly muscles and glands). "Intent" is an interesting aspect of the process in such cases.

Knowledge of this inner body mechanism loop closure led to the desire to find a means to resource biosignals to evaluate the system in use, fine tuning / adaption, and its suitability for each individual.

Systems—From Original Concept Evaluation, to Bespoke and Commercial Apparatus

Biosignals Figure 43.1 illustrates the selected Mindpeak Waverider biological data acquisition system (plug-in sensors not shown) for viewing, archiving and evaluation of the intervention data.

This system was selected because, as well as being able to detect biological signals, the data could easily be mapped as a MIDI protocol to digital synthesizers / DAWs / soundcards, and so forth. This enables a person's various inner electrical biosignals to control digital content, for example to play music. In addition to the auditory stimulus, via MIDI hardware or software translators, the signals can control visual stimuli, such as video games, animations, lighting, and robotic devices. The internal body-sourced data indicates change in electric current produced by the sum of an electrical potential difference across a specialized tissue, organ, or cell system (e.g. the nervous system). In addition to mapping, the MIDI information is scalable and otherwise manipulated in software—Opcode Systems Max was selected at the time (now under Cycling74.com).

Biological data signaling is obtained from electrodes attached to the body, for example on the head / scalp at specific locations to source via an electrophysiological monitoring method to record electrical activity of the brain—commonly referred to as (electroencephalogram recordings—EEG); on the chest area (also further on limb pulse points) to record the electrical activity of the heart over a period of time where electrodes detect the tiny electrical changes on the skin that arise from the heart muscle depolarizing during each heartbeat (electrocardiography—ECG or EKG); and on muscles and muscle groups, where the electrodes detect the

Figure 43.1 WaveRider Pro—four channels physiological signals: brain, heart, muscle signals + GSR. *Source:* © A. L. Brooks.

electrical potential generated by muscle cells when these cells are electrically or neurologically activated. In this body of work a focus was on arm movement so that the signals could be analyzed to detect activation—with conscious intent or unconsciously (e.g. via spasm)—and level of activation (electromyogram or electromyography—EMG). Note that other sensors can be used with muscle and muscle groups where low-frequency vibration may be desirable to detect "mechanical" signal measurements from the surface of a muscle when it is contracted, using either an accelerometer or a microphone placed on the skin over the center of the muscle. At the onset of a muscle contraction, gross changes are evident in the muscle shape causing a large peak in the signal. Following this initial "spike," subsequent vibrations result from oscillations of the muscle fibers at the resonance frequency of the muscle (mechanomyogram—MMG). This technique has a higher SNR, so it is optimal for deeper muscle-activity analysis without the need for invasive methods; forehead adjacent to eye to detect eye movement (electrooculography—EOG); fingers, palms, soles of the feet to detect autonomic nervous system activity via the human's electrically active skin. Change, evident through increased sweating, in electrical conductance between two closely located points over time, signals emotional / sympathetic response and state (galvanic skin response—GSR, or electrodermal activity—EDA). Detection involves sending a small amount of current through the surface of the skin between the points and as sweat contains water and electrolytes, which increase electrical conductivity, this lowers the electrical resistance of the skin.

The selected system also permitted biological data to be mapped to provide auditory digital content: musical notes, scales and controllers (filters, effects, etc.), and visual digital content, such as color filters, shapes, and other attributes, on a variety of contemporary MIDI-compatible hardware and software. Such mapping can include robotic devices (Brooks, 2004d, 2006).

Due to technical advances, MIDI no longer requires cables to operate, as wireless Bluetooth-capable devices are becoming widely used—see the Midi Association web site on these advances (http://www.midi.org).

Audiovisual-related performances were thus attainable and the author added the system to his stage performance art repertoire—thus also controlling stage lighting, scenography, animations, images, ambience, and so forth, as well as auditory collage and improvisations. Notably, there is more to performance and composition in this context than passively allowing the body to generate signals. Each piece demands decisions on what is mapped and precisely controlled—for example, in the audio domain, MIDI scaling and ranges; ADSR envelopes (attack-decay-sustain-release); effects and more; in the visual domain colors, mixes, effects, and so forth. A balance between intent (controlled feed-forward actions) and nonintent (noncontrol) is an innate aspect of performance, in the author's interpretation, giving insight to the healthcare aspects of use. Performance art pieces that specifically targeted noncontrol, where the author created dense information spaces that challenged control from feed-forward conscious thought or in response to feedback stimuli, culminated in a series of live performances titled "Behind the Wall" at the Aarhus Festival Fringe in 1998.

The same gesture-based, gesture-sensitive system was central to the author's international Museum for Modern Art exhibitions and at leading international venues—including the Olympics and Paralympics (1996 and 2000) and the European City of Culture (1996 and 2000). Observations of participants, who had not been instructed how to perform, gives added value as one can analyze the various responses to empowerment. It was particularly interesting to observe groups of children in the interactive installations, as learning and leadership aspects became apparent. Typically, at such venues as museums, Monday was a "dark day"—so that the administration staff was at work but the venue was closed to the public. Agreements were made for such days when the installation was empty so that workshops and access to the installation could be arranged with the author present for groups of impaired persons. So, for example, 20 people in wheelchairs could attend without being hampered by non-wheelchair users, and in such cases a lot of fun and adventure within the installations was evident. Such motivated engagement is a core aspect of the work (see for example the section on the emergent model that developed from the work titled ZOOM—Zone of Motivated Motivation).

Each art-related iteration of the work—stage performances, installations, workshops and showcases—gives insight into the system's evolution and its use across contexts, primarily its use in healthcare. Similarly, use in healthcare gave insight to use in art-related performance. Thus, the term "performance" is used in reference to *human performance* be it under the title of stage artist, installation art (where the audience performs), or in healthcare (supplementing traditional therapeutic intervention in a treatment program where a patient's human performance is in focus). More than three decades of research has shown that it is problematic for the traditionalists to accept this open interpretation. This is true of both the art and healthcare professions, many of whose members over the years played down the achievements and use of such work if it did not fit with their understanding and particular intervention strategy and interpretation of what "performance" referred to. Many actually refused to recognize the work.

Despite this, a large network of interested intellectuals and practicing parties grew from the many displays of the concept, which in turn led to increased dissemination opportunities to learn from the work, giving it momentum. When video games were introduced into the work around 1998, with MIDI mapping to Macromedia (now Adobe) Flash animations, it was heard through the grapevine that professionals in the healthcare sector laughed at their use in therapeutic intervention. In retrospect it could have been easy to give up on the concept but the results were too clear, and this is evidenced by the increased contemporary HCI interest and the many other communities that are now associated with the work and have adopted the concept.

Retrospectively, it is important to state that control and noncontrol performances proved to be a critical learning platform for the commercial invention and development that resulted from the work. However, in the rehabilitation situations that were researched, limitations were evident in biofeedback system use. These limitations included the need to consider patient comfort and how EEG requires an electrode gel applied between an electrode and the participant's scalp to improve the signal quality and reduce noise (improving the signal-to-noise ratio—SNR). Recent advances in EEG electrodes and systems use dry sensors (although some literature suggests that these are inferior to wet sensing). Biosignals also have differing spiking profiles (i.e. latencies), where correlating signals was (and still is) a challenge. At the extremes, intent to activate a muscle can result in an EMG spike that is almost instantaneous, whereas a GSR/EDA exhibits a gradual increase or decrease. Experience from sessions also indicated how the environment setup had to be optimized as much as possible before client entry because many with profound impairment would disassociate if it took too long to start the session because of calibration / optimization. Some even fell asleep while setting up was taking place.

Realizing the need to correlate physiological data with observed (external) behavior and activity, contemporary systems offer synchronized solutions that can be coded according to requirements for observer analysis toward agreed outcomes. In other words, data is archived simultaneously from the different biosignals; however, the variance in lag of indicated effect needs to be considered in analysis. These advanced techniques are not elaborated here. It is also important that creative use and self-analysis of use (i.e. where facilitators try the system themselves) directs interventions, as innate system knowledge and possibilities, as well as limitations, need to be known in order to optimize the solutions.

Sensor-based systems to source beyond inner biosignals—positioning the work Human movement tracking systems can be classified as inside-in, inside-out, and outside-in systems (http://xspasm.com/x/sfu/vmi/publist.html). Each sensor utilized has a distinct profile, and thus innate strengths and weaknesses that need to be considered according to context, use, and desired outcome. Mulder's classification from 1994 is generally still valid in discussing sensor-based systems:

> Inside-in systems are defined as those that employ sensor(s) and source(s) that are both on the body (e.g. a glove with piezo-resistive flex sensors). The sensors generally have small form factors and are therefore especially suitable for tracking small body parts. Whilst these systems allow for capture of any body movement and allow for an unlimited workspace, they are also considered obtrusive and generally do not provide 3D world-based information.
>
> Inside-out systems employ sensor(s) on the body that sense artificial external source(s) (e.g. a coil moving in a externally generated electromagnetic field), or natural external source(s) (e.g. a mechanical head tracker using a wall or ceiling as a reference or an accelerometer moving in the earth's gravitational field). Although these systems provide 3D world-based information, their workspace and accuracy is generally limited due to use of the external source and their formfactor restricts use to medium and larger sized bodyparts.
>
> Outside-in systems employ an external sensor that senses artificial source(s) or marker(s) on the body, e.g. an electro-optical system that tracks reflective markers, or natural source(s) on the body (e.g. a videocamera based system that tracks the pupil and cornea). These systems generally suffer from occlusion, and a limited workspace, but they are considered the least obtrusive. Due to the occlusion it is hard or impossible to track small bodyparts unless the workspace is severely restricted (e.g. eye movement tracking systems). The optical or image based systems require sophisticated hardware and software and are therefore expensive (Mulder, 1994, p. 1).

Over two decades subsequent to Mulder's statement, one can consider how advances in sensing technologies, processing, and miniaturizations (i.e. digital tools both hardware and software) have impacted the field. As with many sensor-based systems, provided the knowledge gained from experience includes knowledge of innate constraints and challenges, one can design HCI-based intervention considering system strengths and weaknesses in line with a mix-and-match strategy.

Figure 43.2 illustrates part of the author's original three-headed sensing system from the 1990s, which was created with infrared sensors that had a volumetric profile in an attempt to move beyond using only worn biometric sensors. System modularity (generally, any number of sensors could be used) was an important aspect to address limitations of the sensing technology profile. However, the technologies used in sensors have limitations and constraints, so choice is dependent on use.

MIDI is an open-signal protocol that doesn't care what the sensing profile is (see Figure 43.3 where three sensing profiles are shown); thus, various apparatus used to sense human input could be mapped within the same software to affect the same or differing content. Software is able to adjust the data so that the effect is of an

Figure 43.2 The author's bespoke three-headed sensing apparatus—version ii left and iii right. *Source:* © A. L. Brooks.

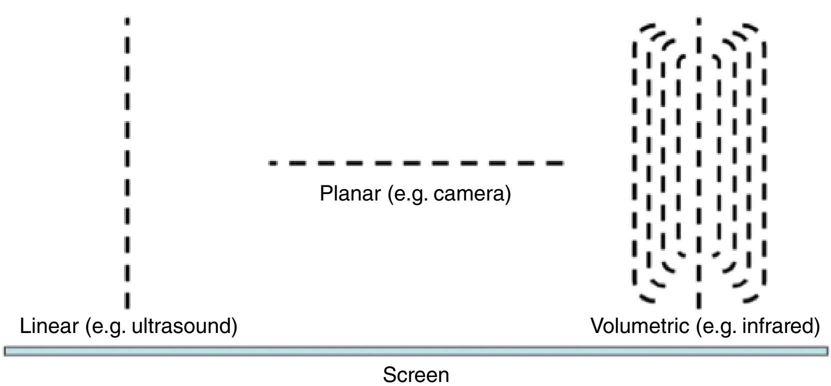


Figure 43.3 Three examples of sensing profiles used in the author's SoundScapes research. *Source:* © A. L. Brooks.

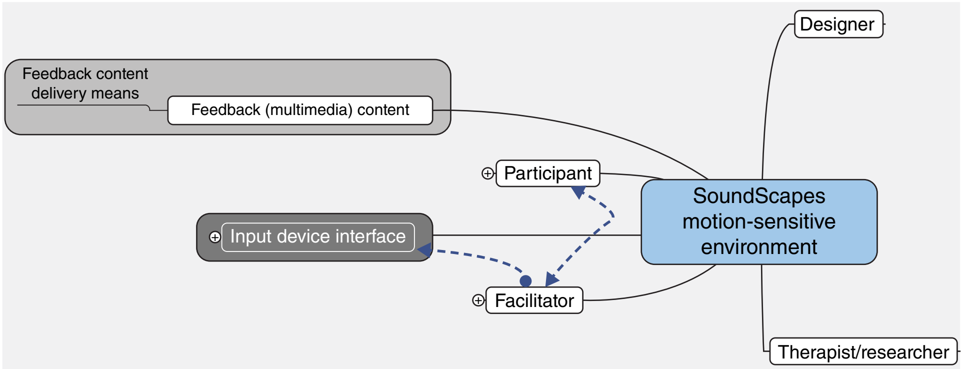


Figure 43.4 Baseline session elements including iterative pre/post designer, therapist, researcher codecisions for achieving the targeted optimal personalization, engagement and motivation for microdevelopment progress. *Source:* © A. L. Brooks.

authoring tool interface that enables tailoring to each individual. MIDI, though a dated protocol, is robust and efficient, and fast enough for HCI.

Parameter changes in such sensor-based HCI systems can take place before or between sessions ("on action") as well as in sessions ("in action"). On-action parameter-change decisions are typically reached jointly, with the medical staff and consulting design team collaborating to determine the presets available as session changes. This strategy enables fine tuning over a series of sessions within a treatment program. Such changes in sessions (in action) involve a process of stepping through predefined session presets that incrementally challenge and stimulate the participant to reach optimal motivation and engagement. In-action decisions are typically decided by the facilitator conducting the intervention; thus, changing the available presets defined beforehand (on action) in the actual session without losing "contact" with the participant is intuitive and based on experience to match session activity to participant profile and progress in order to increase challenges. See below for more on this strategy.

A baseline system, illustrated in Figure 43.4, represents what is typically exposed to participants as the session environment. The collaborating team with consulting designer and the client facilitator develop expected calibrations and profiles prior to initial exposure sessions. The team's knowledge of the client/participant is imperative in this so that role playing can enable an individually tuned environment to be prepared considering targeted innate attributes (Figure 43.5). Ideally such an environment, even in first exposure sessions, if prepared carefully with knowledge (tacit and beyond), skill, and competence, enables a participant to enter and begin with minimal recalibration to maximize motivation for all (not just the client or participant, but also the facilitator experience is affected—as in Brooks, 2005, 2010, 2014).

Emergent model: Zone of optimized motivation (ZOOM)

In the context of the research presented in this chapter, the requirement for ongoing assessment within sessions (i.e. *in action*) is integral to the adaptive design concept to optimize iterative tuning of intervention, including system changes depending on

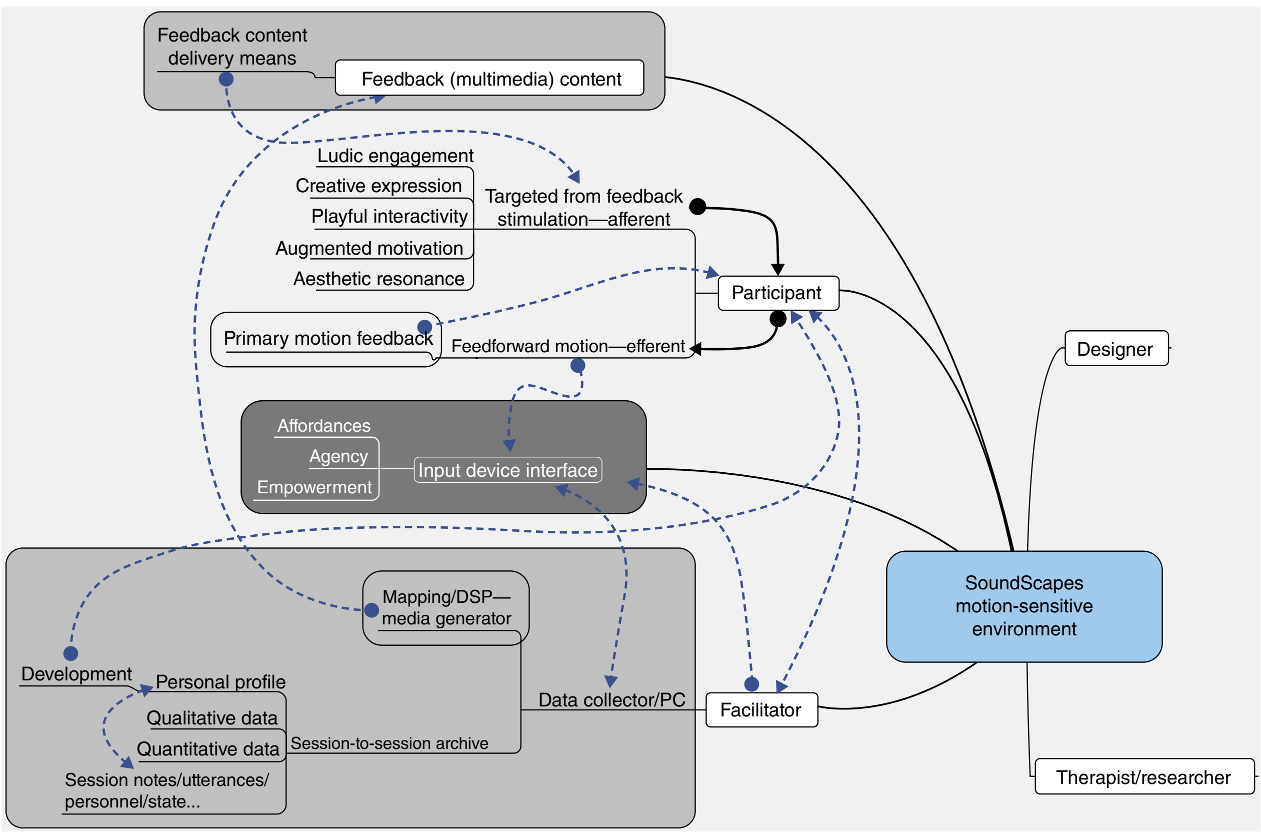


Figure 43.5 Elaborating previous Figure 43.4 to illustrate additional components—complexity and interrelationships of data capture aligned to goals—not presented to participant or facilitator. *Source:* © A. L. Brooks.

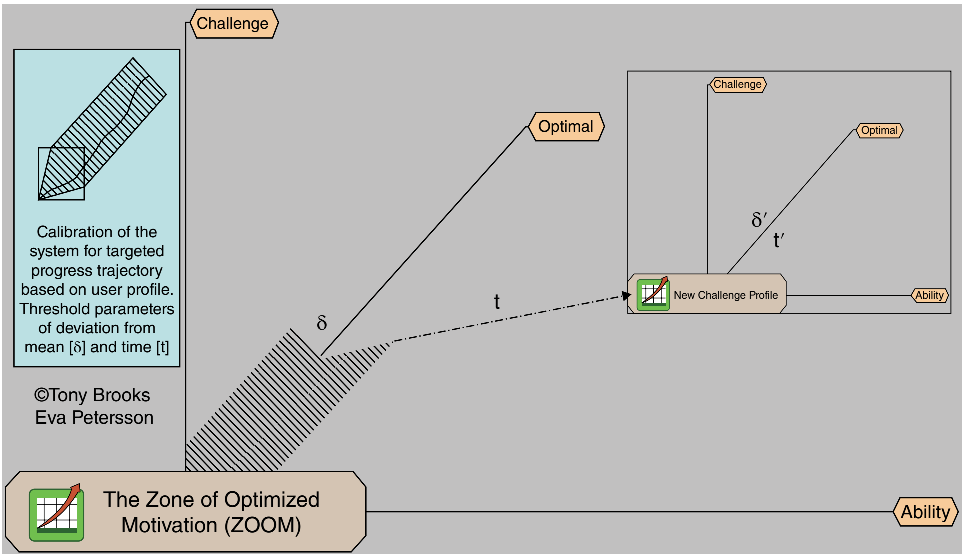


Figure 43.6 ZOOM—The Zone of Optimized Motivation (version 1): Emergent in-action intervention model as synthesized from key theories including microdevelopment (e.g. Fischer), in-action / on-action (Schön), Flow (Csikszentmihalyi)…A model for fine-tuning system change parameters in live work. *Source:* © A. L. Brooks.

progress and participant engagement, enjoyment, and flow state. This *in-action* model is titled the zone of optimized motivation (ZOOM—see Figure 43.6); it complements an *on-action* evaluation strategy, based upon recursive reflection, which critiques and reflects postsession on the overall treatment program as well as each session and possible improvements that can be implemented. The two aligned models (see Brooks, 2005, 2011) evolved with the author's doctorate research, alongside apparatus development and software integration, building upon his prior research, which realized a patent titled "Communication Method and Apparatus" (US 6893407—original document from 2000).

Whilst the hardware apparatus (e.g. sensing devices) has strengths and weaknesses, which should be considered in the design phase of a treatment program, it is equally important to consider digital responsive content, which similarly has constraints and challenges to consider. So far, in this chapter, data sourcing has been presented with mapping to primarily audio content. The following statement exemplifies linkage between music and body:

> Whatever else music is about, it is inevitably about the body: it is invariably an embodied practice. When we hear a musical performance, we don't just "think," we don't just "hear," we participate with our whole bodies. We enact it. We feel melodies in our muscles as much as we process them in our brains—or perhaps more accurately, our brains process them as melodies only to the extent that our corporeal schemata render that possible (Bowman, 2000, p. 50).

However, during the 1990s the mapping to auditory stimulus was supplemented to realize a multimodal stimuli environment so that additional selection and system tuning was available.

Beyond sound

Experiences from the research exploring interactive audio made clear how some clients had an individual preference for other modalities of stimuli rather than sound alone. Such a preference was evident from early studies conducted by the author on visual stimulus in the form of analog video feedback (http://softology.com.au/videofeedback/videofeedback.htm). Such analog video feedback, as illustrated with the upper torso in Figure 43.7 and hand images in Figure 43.8, used the mirroring of input gestures to engage participants into "their own" creative environment. This technique relates but differs from physiotherapy intervention in using a traditional body length silver mirror as evidenced in the author's empirical work where a patient was mirrored seeing his own body reflected to strengthen associations such as proprioception, concentration, eye-hand coordination, and self-agency.

To achieve this form of analog video feedback, a patient is placed in a position within a specific system setup (video camera plus feedback monitor) to be able visually to sense direct associations, effects, and consequentially control, from interacting with a responsive visual stimulus that is consequential to a conscious or unconscious intent action—i.e. feed-forward input creating abstract responses and thus interactions that stimulate the participant to move.

A variety of techniques and add ons can be employed to vary the interactions in order to maintain engagement in such an analog video mirroring session. The author's early research and experimentation with this analog technique from the 1980s was developed as an ongoing supplement to the developing interactive "system."

For example, Figure 43.7 illustrates a "private space" in one of the author's annual fortnight residential workshops held at Casa da Musica, Porto, Portugal, where each

Figure 43.7 A participant with impairment interacts with responsive analogue and digital media in the author's 2008 Casa da Musica, Porto, Portugal session—organized as an annual two-week workshop. *Source:* © A. L. Brooks.

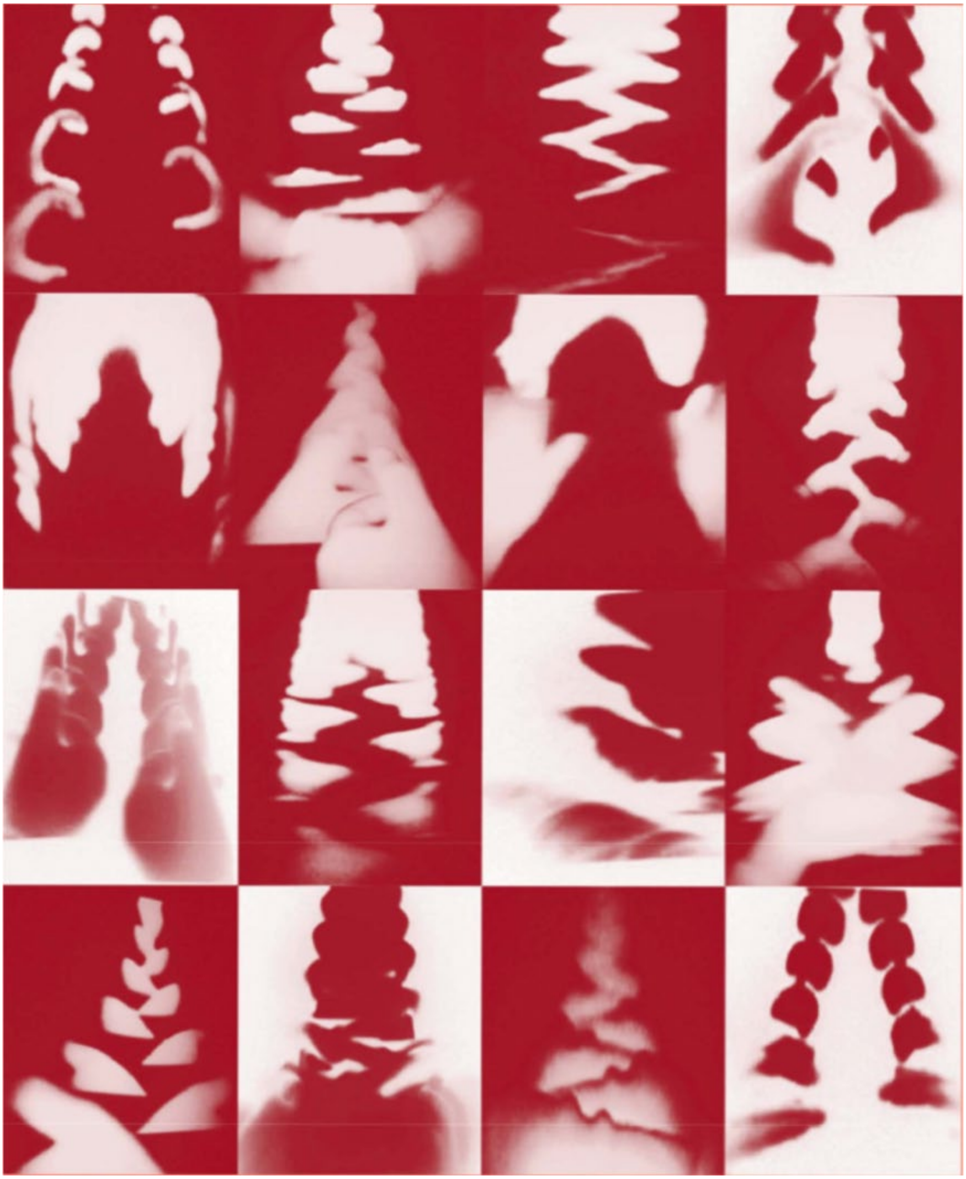


Figure 43.8 Example compilation images of analogue video feedback experiments in (re) habilitation session (hands—single and double participants from EU project CAREHERE project), which resulted from the research. *Source:* © A. L. Brooks.

day two morning groups and two afternoon groups of approximately 20 participants at a time would attend. The setup involved analog video feedback, plus RGB lighting and a microphone that was mapped to change colors. These workshops are subject of a series of DVD publications.

Figure 43.8 was a part of the CAREHERE (http://www.bristol.ac.uk/carehere/) European project funded under Framework V IST Key Action 1 supporting the program for Applications Relating to Persons with Special Needs Including the Disabled and Elderly, which the author coordinated when based in Sweden.

Such was the success of the analogue video feedback research, indicating how direct visual feedback that mirrored human feed-forward input could be used, that research began on using responsive robotic lighting devices (circa 1994) and later also the Flash software program (then Macromedia, now Adobe) to create gesture-controlled interactive video games (circa 1998).

Unencumbered gesture control of robotic lighting devices

The use of lighting devices began by mapping MIDI data sourced from body gestures to Martin Light units. Martin is a Danish company whose headquarters (then in Risskov) were located close to where the author lived in Aarhus, Denmark's second city (this being where the subsequent Martin headquarters are located).

A sponsorship was agreed and covered two phases (a) moving mirror lighting units and smoke machine, and (b) state-of-the-art moving head lighting units as a multiple of three to match the typical sensor device set up where each sensor could align with a distinct lighting-control channel (see Figure 43.9) as elaborated in Brooks (2004d).

To enable control of lighting units, the author purchased an Elektralite CP10 lighting programming rack device in 1995 to translate MIDI-to-DMX. DMX512 (Digital Multiplex) is the "language" understood by lighting devices to enable programming via a specific unit address (Adr) from control devices. With a MIDI-to-DMX512 translator, robotic device control (and in this case lighting) was enabled from sensed human motion under the umbrella of the author's research. An element

Figure 43.9 MiniMac lighting unit with segmented detail of mappings / control via Elektralite CP10 MIDI/DMX 512 signal conversion interface. Upper left illustrated programming window, lower left gobo wheel, upper right transport, and lower right lighting color. *Source:* Image compilation © A. L. Brooks—source images used with permission Martin Lights, Denmark.

of the control of lighting by motion links to how a performer's movement on a stage can be automatically tracked via MIDI. MIDI Show Control (MSC) is a real-time system-exclusive extension of MIDI that enables devices to communicate with each other and with computers in order to perform control functions in live and canned entertainment applications. In this context the author used the Cycling74 software Max upstream of the MIDI-to-DMX translator to tailor gesture control to each individual (Brooks, 2002).

In addition to the use in rehabilitation research, the control of lighting units was showcased at both of the author's productions for the cultural event that supported the Olympics / Paralympics in 1996 (Atlanta) and 2000 (Sydney). Notably, both of these events also had an associated scientific conference where the concept was also presented under healthcare and rehabilitation. The concept was also presented at other major events around the world. The author coordinated with Danish lighting hardware manufacturers, Martin, who sponsored the research, to enable use of local hardware (from their national distributors) to save transportation costs (see the *Four Senses* TV documentary at https://www.youtube.com/watch?v=gTjvCh-XB2o).

Unencumbered gesture control of video games

Following the successes of gesture control of robotic devices as a HCI-focused intervention for healthcare and rehabilitation, personal computer-based video games were introduced. A basic flying plane game was the first where an unencumbered gesture could cause an animated plane to take off and maneuver and then land. It was the first gesture control of a game.

Unencumbered gesture-controlled interactive games enhanced the SoundScapes therapeutic intervention environment around 1998 by introducing competitive challenge in the research. It was also conceived that introducing video games would improve options and potentials for facilitators, as well as making the sessions more fun through an increase in social engagement.

It was clear, in studies, that competition in the form of challenging users with their own scores, and increasing the levels of challenge, motivated participants to play the games, but the therapists / facilitators involved reportedly found it difficult to adopt games into their interventions. This was due to various factors, such as the stigma associated with playing video games (the belief that they were for children and not a serious tool / aid for supporting therapeutic intervention), a lack of control over the situation, a belief that technology was possibly taking their jobs, and technophobia.

The evaluations from the studies that included a reported desire for more control by therapists / facilitators led to an interface for easily adapting each game to be tailored to each individual. Figure 43.10 illustrates the author presenting a Flash-based game that was developed to supplement acquired brain injury rehabilitation (three of the author-created "i3 light gobos" are also visible). The game environment is seen on the right screen—underwater with wireframe dolphin catching fish—time, score, and level details are included—see also Figure 43.11. On the left screen is the developed tailoring interface where game parameters can be changed and activity archived.

Figure 43.10 Gesture-based wireframe Dolphin game at Orbit-Comdex i3 village installation, 2001: Left screen shows the authoring / personalization tool where changes are implemented—right screen shows game. *Source:* © A. L. Brooks.

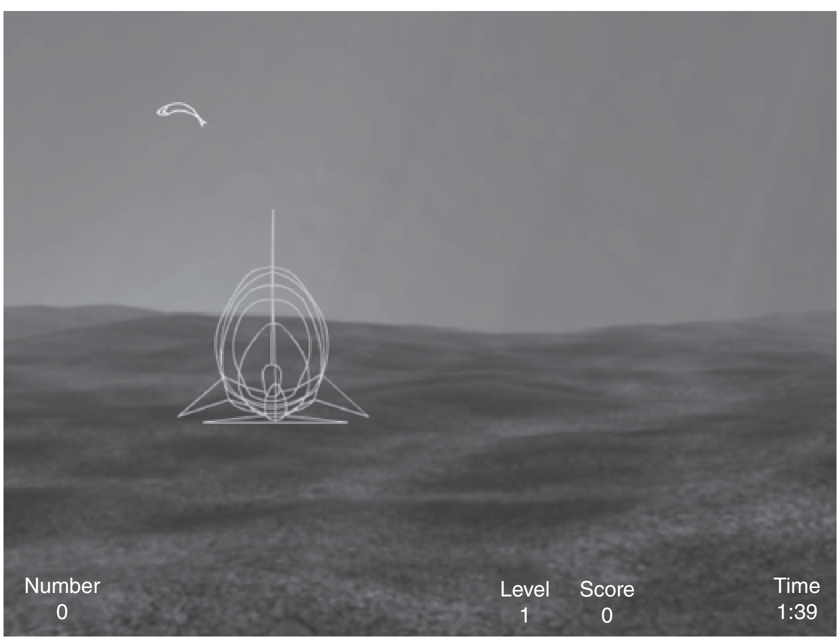


Figure 43.11 Dolphin Game—see also Figure 43.10—illustrating time, score, level, and number of fish caught—quantifiable session data resulting in treatment programme competitive / motivated participants. *Source:* © A. L. Brooks.

The intervention using this game was designed so that a therapist would control one sensor mapped to horizontal travel of the dolphin, and the patient would control a second sensor mapped to the vertical control of the dolphin. This involvement of the physiotherapist sharing the game controls gave added support and value to the environment and was clearly high in social interaction as each party worked with the other to achieve the challenges. Progress in the game was through various levels and using various sides of the body / various arms (affected and unaffected). However, at a point that the therapist considered appropriate, the patient took control of both sensors to work both arms (affected and unaffected) together.

The above strategy of use utilized "improvisation" by the facilitator to match and optimize the experiences of the participant. This is also a key factor in the need to educate therapists and others in healthcare about the potential of this approach when appropriate, as opposed to the facilitator being restricted because of rules and delimitations.

Human-computer interaction rehabilitation and healthcare sessions in the author's work focuses on creating a situation where the "response to intent is so immediate and aesthetically pleasing as to make one forget the physical movement (and often effort) involved in the conveying of the intention" (Brooks, 2003, 2004a, 2004b, 2004c, 2004d; Brooks et al., 2002). This situation is linked to our human Kinesphere (see Laban, 1963, p. 85).

Figure 43.12 depicts a participant in a Kinesphere. Gestures are captured and data is mapped to control selectable multimedia feedback determined by participant profile (preferences, needs, and desires) and the therapists' goal for progress intervention strategy. Such adoption of intervention strategies using digital content is widely reported, especially with video games—due to the influx of affordable natural interfaces / motion controllers.

Researchers from the United States and the United Kingdom, in visits to Aarhus in 2001, saw the SoundScapes system (under the name of Personics) being evaluated positively for its potential in the field. Such evaluation supplemented those from the expert healthcare evaluations already obtained.

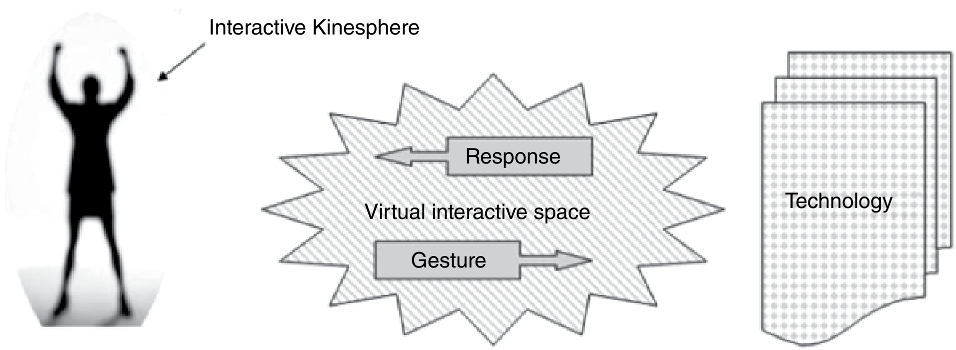


Figure 43.12 Linking Laban's Kinesphere to the SoundScapes interactive Design space. *Source:* © A. L. Brooks.

Conclusions

In addition to the experiences and tacit knowledge obtained from exposure to family members having profound disabilities, the author's background included employment (1980–1986) involving large mainframe distributed control system (DCS) computers (i.e. Honeywell TDC-2000) as well as personal computers at home. UNIX and its companion networking technology TCP-IP were utilized alongside Ethernet in the DCS systems. In addition to experiences with computers, art, especially music, was prevalent in the family home and the author showcased at the Institute of Contemporary Arts (https://www.ica.org.uk) in 1978 at age 22. Other noteworthy spaces and events have since presented his work, for example museums of modern art, the Olympics/Paralympics, European City of Culture, Danish NeWave in New York, and others. Together, these combined experiences resulted in the author conceiving the bespoke systems researched—including the patented communication apparatus and method. Concepts of note include the e-health SoundScapes system researched under the Danish government-funded Humanics project program (1996–2002) appendix 1.

This chapter discussed computer programming to complement tacit knowledge and the possibilities of learning from those with impairment to advance the field. Thus, the design of the bespoke systems created (e.g. see Brooks, 1999) as well as the early adoption of commercial video game controllers that were based on gesture (held, worn, or free standing—commonly referred to as a natural user interface—e.g. Sony EyeToy, Playstation Eye, PlayStation Move; Nintendo Wii Remote, Nunchuk and balance board; and the Microsoft Kinect). Of these major producers, i.e. Sony, Nintendo and Microsoft, only the latter openly shared development kits to enable third-party coding. In the case of the Wii devices, additional commercial software called OSCulator (http://www.osculator.net) enables mapping of sourced data. This software also enables mapping of other Nintendo Wiimote extensions: Motion Plus, Guitar Hero World Tour Guitar and Drums, and the Classic Controller. It also enables control from the Wacom Tablet; 3Dconnexion's SpaceNavigator; traditional mouse and keyboard; TUIO open framework protocol, and API tangible multitouch surfaces such as reacTIVision. It also features advanced OSC routing and bidirectional MIDI with TouchOSC. Mobile devices (Mobile Apps on iOS, Android) can also be used. Such access for developers to be able to map controllers to content is imperative and something sadly missing from the original Beamz laser system, which, up until its 2015 model (C1R42; black with red lit logo) only allowed access to the company's proprietary content (the C1R42 enables only MIDI note triggering).

With so many used game controllers that would require minimal investment to purchase, such as the Nintendo Wii, there is a world of HCI opportunities in therapeutic environments if staff are motivated to explore its potential for their clients. It is not a financial investment but one of time and interest that can benefit all.

Enculturation, in the context of this work, is the process by which one learns and comprehends significant aspects of the partner organization—for example, framework, concept, direction, strategy, and requirements of the surrounding culture—to best assimilate its practices and values and so optimize activities and thus outcomes. Supporting the enculturation of the researcher is a core responsibility that leaders should assign to staff at the organization that the HCI/SoundScapes researcher/designer attends to collaborate on a project. Unfortunately, this aspect is sadly missing, and the subject of researcher enculturation and alignment to policies and personnel is rarely listed in project applications or collaboration agreements.

Coda

The mixed methods approach to qualitative and quantitative data collection and analysis enables transdisciplinary discussions on outcomes. In line with this, emergent models are worthy of study and further research.

By including complementary responsive and interactive content in video games, for example, with creative expression through digital music making, painting, and robotic control, more options for engagement become apparent. Aesthetic Resonance is also posited as an area for further development research (see Brooks, 2005, 2011; Brooks et al., 2002). Changing content has been supported as a means to increase engagement and achievement whilst optimizing motivation and fun for all involved (i.e. participant and facilitator). There is clear evidence that these approaches can enhance creative rehabilitation interventions, session compliance, and participation. This is further evinced, for example, by an independent third-party therapist study with outcomes reporting up to 400% improvement in training-specific performance compared to traditional intervention with balance training for the elderly. In the study, elderly frail patients were able to increase muscle strength and physical endurance, alongside improved static balance results so that clinical rehabilitation impact was evaluated (Hagedorn & Holm, 2010).

Future work

Research has been located in a virtual reality—interaction design—human behavior analysis laboratory complex at the Esbjerg campus of Aalborg University in Denmark[1,2]. In this environment a 2 m high × 5 m wide back-projected screen, optimized for projections, has been used with sensing of head tracking and participant location. The laboratory was dismantled in 2013 and since then research has been focused on mobile solutions, especially head-mounted displays (HMDs), such as Oculus Rift, HTC Vive, HDK OSVR, etc.

Future work includes improving the sourcing of quantitative data relating to participant experience within the designed environments with interaction options. Recently 'in-HMD' brainwave (https://myndplay.com) and tracking/dilation (https://pupil-labs.com) devices have been explored so that physiological data can be gathered about the user's experience to provide information to the environment designer to allow the iterative tailoring/personalization of challenges as outlined above.

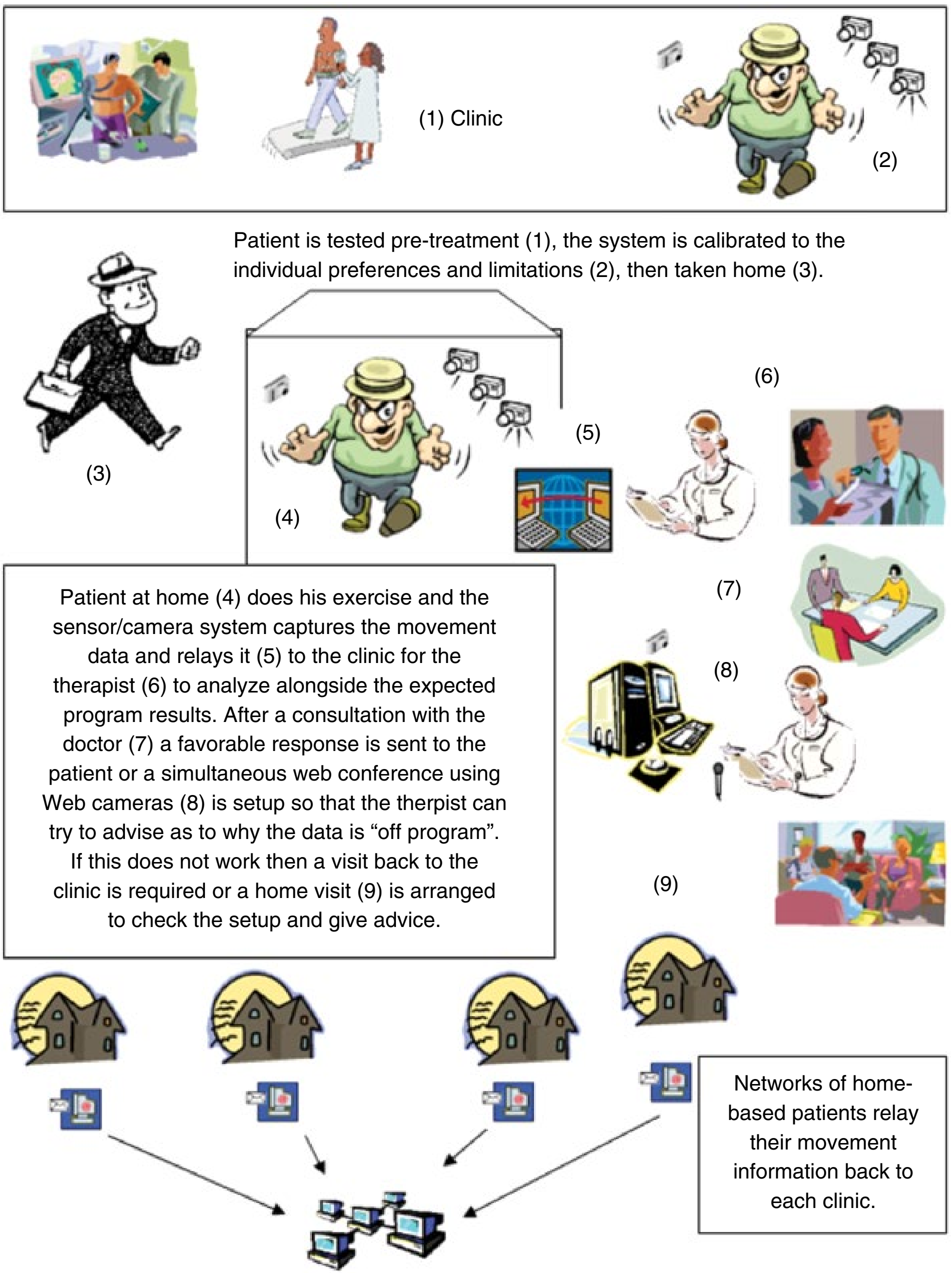
The research also targets the implementation of ICT to improve the quality and ease of use of assessment-testing tools while improving the assessment-testing tools' reliability and validity. For example, targeting improved "quality" in psychometric terms relates to reliability, measurement error, temporal stability, sensitivity, specificity, predictive validity, and the care with which test items are derived and normative data obtained (e.g. see Slick, 2006).

Acknowledgements

Current and previous sponsorship of the SoundScapes body of research is acknowledged from Marshall Amplification, UK; IBM Denmark; Martin Lights, Denmark; Hornberg Research; IK Multimedia; Livid Instruments, and TC Electronics.

Certain images and sections of this text constitute three decades of research, and thus have been presented elsewhere and in the author's PhD thesis. In stating this, the content of this chapter is original in its formulation and compilation.

Appendix 1: 'Humanics project' design proposal for Acquired Brain Injured (ABI) patient home training using Internet with clinician remote monitoring (c. 1996)



(© A. L. Brooks)

End Notes

1 http://isvr.org/wp-content/uploads/ISVR-Newsletter-Issue7-2016-04.pdf
2 http://www.openscenegraph.org/index.php/gallery/use-cases/90-sensoramalab

References*

Bowman, W. (2000). A somatic, "Here and Now" semantic: Music, body, and self. *Bulletin of the Council for Research in Music Education*, *144*, 45–60.

Brooks, A. L. (1999). Virtual interactive space (VIS) as a movement capture interface. In E. Miller & L. Miller (Eds.), *Proceedings of the International Symposium on Integrative Medicine and Expressive Therapies* (p. 3). Rhinebeck, NY: Omega Institute.

Brooks, A. L. (2002). "SoundScapes" as a human computer interface for use in therapy, elderly care and rehabilitation. In *Proceedings of the Fifth Asia Pacific Conference on Computer Human Interaction: User Interaction Technology in the 21st Century* (pp. 833–844). Beijing: Science Press.

Brooks, A. L. (2003). Soundscapes: Multisensory reciprocity through subliminal non-control. In *Proceedings of the Sixth International Research Conference: Consciousness Reframed: Qi and Complexity* (pp. 37–44). Beijing: University of Beijing.

Brooks, A. L. (2004a). Body electric and reality feedback loops: Virtual interactive space and entertainment. In Proceedings of the 14th International Conference on Artificial Reality and Telexistence (ICAT 2004) (pp. 93–98). Seoul: Advanced Institute of Science and Technology.

Brooks, A. L. (2004b). Creating aesthetic resonant environments for the handicapped, elderly and rehabilitation: Sweden. In *Proceedings of the Fifth International Conference for Disability, Virtual Reality and Associate Technologies (ICDVRAT2004)* (pp. 191–198). Reading: University of Reading.

Brooks, A. L. (2004c). Interactive painting: An evolving study to facilitate reduced exclusion from classical music concerts for the deaf community. In *Proceedings of the Fifth International Conference for Disability, Virtual Reality and Associate Technologies (ICDVRAT2004)* (pp. 101–11). Reading: University of Reading.

Brooks, A. L. (2004d). Robotic synchronized to human gesture as a virtual coach in (re)habilitation therapy. In *Proceedings of the Third International Workshop on Virtual Rehabilitation (IWVR2004)* (pp. 17–26). Lausanne: VRlab, École Polytechnique Fédérale de Lausanne (EPFL).

Brooks, A. L. (2005). Enhanced gesture capture in virtual interactive space (VIS). *Digital Creativity*, *16*(1), 43–53.

Brooks, A. L. (2006). SoundScapes—Beyond interaction: In search of the ultimate human-centred interface. In *Proceedings of the 16th International Conference on Artificial Reality and Telexistence (ICAT2006)* (p. 17). Zhejiang, China: Zhejiang University of Technology.

Brooks, A. L. (2011). *SoundScapes: The evolution of a concept, apparatus, and method where engagement in virtual interactive space is a supplemental tool for therapeutic motivation* (Doctoral dissertation). Institut for Arkitektur og Medieteknologi, Aalborg, Denmark.

Brooks, A. L. (2014). Camera-based software in tehabilitation/therapy intervention. *Journal of Accessibility and Design for All*, *4*(2), 130–143.

Brooks, A. L., Hasselblad, S., Camurri, A., & Canagarajah, N. (2002). Interaction with shapes and sounds as a therapy for special needs and rehabilitation. In *Proceedings of the Fourth International Conference on Disability, Virtual Reality and Associated Technologies (ICDVRAT2002)* (pp. 205–212). Reading: ICDVR AT/University of Reading.

* The author's works were originally published using slightly different forms of his name, such as T. Brooks, A. L. Brooks, and Lewis-Brooks. A. L. Brooks is used for all of these works here.

Brooks, A. L. & Monção, A. (2010). Art, brain and languages editorial. *International Journal of Arts and Technology*, *3*(2–3), 137–153.

Hagedorn, D. K. & Holm, E. (2010). Effects of traditional physical training and visual computer feedback training in frail elderly patients. A randomized intervention study. *European Journal of Physical and Rehabilitation Medicine*, *46*(2):159–168. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/20485221

Laban, R. (1963). *Modern educational dance* (2nd ed.), London: MacDonald & Evans.

Mulder, A. G. E. (1994). *Human movement tracking technology* (Report), Burnaby, BC, Canada: NSERC Hand Centered Studies of Human Movement Project, Simon Fraser University.

Slick, D. J. (2006). Psychometrics in neuropsychological assessment. In E. Strauss, E. M. S. Sherman & O. Spreen (Eds.), *A compendium of neuropsychological tests* (pp. 3–43). Oxford: Oxford University Press.

44

Visual Analytics for Comparing Multiple Clustering Results of Bioinformatics Data*

Sehi L'Yi, Bongkyung Ko, DongHwa Shin, Young-Joon Cho, Jaeyong Lee, Bohyoung Kim, and Jinwook Seo

Introduction

Since Eisen lab's *Cluster* and *TreeView* (Eisen, Spellman, Brown, & Botstein, 1998) popularized cluster analyses and visualizations of microarray data, cluster analysis has been widely used in the bioinformatics community. As genetic probing technologies rapidly improve in capacity and accuracy (e.g. next-generation sequencing), cluster analysis is playing an even more important role in the descriptive modeling (segmentation or partitioning) of the large data produced by high-throughput probing technologies. Although cluster analysis has become a routine analytic task for bioinformatics research, it is still arduous for a researcher to quantify the quality of a clustering method's clustering results.

There have been a few attempts to develop objective measures for clustering quality assessment; however, in most practical research projects, determining the quality of a clustering result is subjective and application specific (Seo & Shneiderman, 2002). To make things even more challenging, there are a large number of clustering methods, which could generate diverse clustering results. Moreover, even an individual clustering algorithm could end up with different results depending on the clustering parameters.

There is no generally accepted objective metric for selecting the best clustering method and its parameters for a given dataset, so researchers often have to run multiple clustering algorithms and compare different results while examining the concordance/discordance among them. Such a comparison task with multiple clustering results for a large dataset is cognitively demanding and laborious.

*This research was originally published in L'Yi et al. (2015).

The Wiley Handbook of Human Computer Interaction, Volume 2, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.

In this research, we present *XCluSim*, a visual analytics tool that enables users to compare multiple clustering results interactively and explore individual clustering results using dedicated visualizations.

Related Work

Visual comparison using visualizations for multidimensional categorical data

Since multiple clustering results can be treated as multidimensional categorical datasets, they can be visualized using various techniques corresponding to the specific data types. These techniques include *Parallel Sets* (Bendix, Kosara, & Hauser, 2005) and *Parallel Coordinate Plots* (Inselberg & Dimsdale, 1990). Much prior work on the visual comparison of multiple clustering results employed these techniques (Ding, Wang, Huang, & Machiraju, 2014; Havre, Shah, Posse, & Webb-Robertson, 2006; Lex et al., 2012; Lex, Schulz, Streit, Partl, & Schmalstieg, 2011; Lex, Streit, Partl, Kashofer, & Schmalstieg, 2010; Pilhofer, Gribov, & Unwin, 2012; Seo & Shneiderman, 2002; Zhou, Konecni, & Grinstein, 2009). Here we focus our discussion on the ones that are most relevant to us in terms of utilizing ribbonlike bands to represent concordance / discordance among multiple clustering results.

In *iGPSe* (Ding et al., 2014), to visually compare clustering results of two different expression data types (i.e. gene expression and micro-RNAs expression), two dimensional axes were juxtaposed, allowing for the use of parallel sets. By observing the flow of ribbonlike bands, users were easily able to see which items were shared between a pair of clusters from two different clustering results. *HCE* (Seo & Shneiderman, 2002) also juxtaposed a pair of hierarchical clustering results in parallel to enable comparison tasks with the two results. In contrast to *iGPSe*, *HCE* used a partitioned heatmap instead of a simple node to show the details of each data item. To reveal the relations between items in a pair of heatmaps, matching items were connected with straight lines. However, these two visual analytics tools only supported the comparison of a pair of clustering results. Moreover, because they used connectivity between related items, it was often the case that there were too many crossing lines with a large dataset.

CComViz (Zhou et al., 2009) alleviated the line-crossing problem while focusing on the comparison tasks of more than two clustering results. In their work, multiple clustering results were visualized with a parallel coordinate plot: clustering results as dimensions, clusters as vertical positions in each dimension, and items as lines. Users could grasp the overall distribution of items across multiple clustering results by tracking the flow of lines crossing multiple dimensions. Similar representations were used by Havre et al. (2006), but *CComViz* devised an algorithm for rearranging clusters and their members to minimize visual clutter between each dimension.

Matchmaker (Lex et al., 2010) also utilized the parallel coordinate plot, but to show raw data simultaneously, partitioned heatmaps were shown in dimensional axes. The items in each dimension were rearranged by their average values so that heatmaps clearly showed the patterns of the raw data. Unlike the case of *CComViz*, in this case, partitioned heatmaps used a bundling strategy to maintain the position of each item in a dimension. This reduced line crossings between adjacent dimensions. Although this method generated a clearer overview of the distributions of items, it had some drawbacks. First of all, the flows of inner lines were invisible unless users explicitly highlighted the lines. Secondly, since the lines were bundled, the width of a band may not have accurately conveyed the number of the items belonging to the band.

CComViz and *Matchmaker* are probably most relevant to *XCluSim*. They depended on a linear ordering of dimensions (or clustering results), which made it difficult to do all-pair comparisons with a large number of clustering results at once. For example, as the authors said, *Matchmaker* only enabled users to compare, at most, six clustering results simultaneously, even with the limited linear ordering of dimensions. The same dataset can yield a large number of different clustering results, so it is necessary to provide a more scalable way of comparing them. In *XCluSim*, we present diverse overviews to help in comparison tasks with many clustering results.

Visualization using similarity measures

There are a few approaches to visualizing measured similarity values between clusters (or items) in different clustering results instead of explicitly visualizing shared items among multiple clustering results. Sharko et al. (2007) used a color-coded similarity matrix view to show the stability between items or clusters across different clustering results. Similarities were measured by counting how many times each pair of items was clustered together or how many items each pair of clusters shared. Kothur, Sips, Dobslaw, and Dransch (2014) used bar charts arranged in a matrix layout to show similarity values between a pair of clusters. However, these two works were restricted to comparing a pair of clustering results because they both used a matrix layout.

iGPSe (Ding, Wang, Huang, & Machiraju, 2014) used *Silhouette Plot* (Rousseeuw, 1987) to help compare a pair of clustering results. Each item received a standardized dissimilarity value ranging from –1 to 1. This value represented dissimilarity in such a way that, when a value was close to 1, its average dissimilarity from all other items in the same cluster was much smaller than the maximum average dissimilarity from all items in another cluster. When the value was close to –1, the meaning of the value was reversed. By representing these similarity values between clustering results using a bar chart, users were able to assess the relative quality of clustering results.

These previous works using similarity measures allowed for comparisons of only a small number of clustering results. However, it is clear that, by abstracting detailed differences to simpler similarity measures, the visual comparison could be rendered more scalable. In our work, we used a graph layout and a dendrogram to show similarity overviews in a more scalable way.

Color encoding for clusters

Color is a powerful visual cue for representing a cluster membership. It is used in many visualization techniques, including parallel coordinate plots (Sharko et al., 2007; Zhou et al., 2009) and scatterplots (Andrienko et al., 2009; Hossain et al., 2012; Kandogan, 2012), to discriminate clusters while revealing trends in the raw data. Similar efforts exist in the visualizations of multiple clustering results. For example, when using the parallel sets view, a few distinct colors are used to encode each cluster to distinguish it from others (Ding et al., 2014; Zhou et al., 2009).

However, if there are clusters from different clustering results that share the same members, it is not desirable to encode them in distinct colors since it may mislead a user into thinking that those clusters are different. Moreover, when the number of clusters increases, it is hard to color code clusters differently because it is hard to discriminate between more than 10 colors.

A useful color-encoding strategy is *Tree Colors* (Tennekes & de Jonge, 2014), which was devised for tree-structured data to represent similarities between nodes. A part of the parent's hue range is recursively assigned to its child nodes. As a result, nodes with the same parent have similar colors, while those that are less similar have different colors. Moreover, this color scheme reflects the level of a node by using differentially encoded chroma and luminance in each level. If the similarities between clusters from multiple clustering results can be represented as a tree structure, *Tree Colors* may be well suited to represent similarity among them. In *XCluSim*, we used this color scheme to color-code clusters after building a hierarchical structure by running a hierarchical agglomerative clustering (HAC) (Eisen et al., 1998) with all clusters.

Task Analysis and Design Goals

When performing a cluster analysis with a gene expression dataset, bioinformaticians typically follow an iterative analytics process: (a) they filter out unnecessary genes from the dataset for more focused analysis; (b) they run a clustering algorithm with the selected genes; and (c) they validate clusters in the clustering result to determine whether genes are clustered properly in the biological context. When the quality of the clustering result is not satisfactory at the validation stage, they often have to return to previous steps and run the same clustering algorithm with different parameters or run a different clustering algorithm.

Years of close collaboration with bioinformaticians have revealed to us that they often faced challenges in this iterative analytics process. First of all, there is no flexible analytics environment that supports them through the iterative process while providing diverse clustering algorithms and keeping track of their exploration history (i.e., the sequence of the clustering algorithms and parameter settings). Moreover, it is challenging for them to compare effectively different clustering results generated during multiple iterations while investigating the quality of the results at diverse levels (i.e. clustering results level, cluster level, and gene level).

To address these challenges in the iterative process of cluster analysis, we set the following design goals for our visual analytics tool:

- to facilitate scalable visual comparison of many clustering results at diverse levels;
- to support the generation of diverse clustering results;
- to promote understanding of the characteristics of each clustering algorithm and its parameters in results;
- to provide dedicated visualizations effective for different types of individual clustering results.

We designed *XCluSim* based on the *visual information seeking mantra* (i.e. overview first, zoom and filter, and details-on-demand) (Shneiderman, 1996) to support scalable visual comparison better. Since each combination of different clustering algorithms and their parameters may yield different clustering results, it is inevitable from those many clustering results to (a) *see their overall similarity first*, (b) *choose a subset of them*, and then (c) *perform detail comparisons and explore individual clustering results.*

XCluSim provides as many clustering options as possible by implementing famous clustering algorithms and linking the clustering algorithms available in Weka (Hall et al., 2009). It also keeps track of clustering options that users try during the analysis process.

Visualization Design for *XCluSim*

In this section we introduce visualization techniques and user interactions for comparison tasks. They include overview, filtering/selection, and detail view. Then we present visualization techniques that help users to explore individual clustering results. For better comprehension of the visualization components in *XCluSim*, we first describe a color-encoding strategy for clusters, which we consistently apply to every visualization component of *XCluSim* prior to explaining each visualization.

Color encoding of clusters using tree colors

To help users identify similarities among multiple clustering results, we color code each cluster based on *Tree Colors* (Tennekes & de Jonge, 2014), which provides a color-coding scheme for tree-structured data. We first hierarchically cluster all clusters from every clustering result using HAC. The correlation coefficient is used as the similarity measure between a pair of clusters as in Zhou et al. (2009). This maintains consistency in the use of the cluster similarity measure in *XCluSim*, which is also used for rearranging bands (i.e. clusters) in the *enhanced parallel sets view* (see the Enhanced parallel sets view section). In the resulting tree-structured cluster hierarchy, we assign an appropriate color to each cluster based on the *Tree Colors* color-coding scheme so that similar clusters have similar colors.

This color encoding helps users intuitively assess the similarity of clusters. For example, in Figure 44.1(d) (the *enhanced parallel sets view*), ① and ② have very similar colors while ① and ③ do not, which means that ① and ② share most items while ① and ③ barely share any item. This color-coding scheme is consistently applied to overviews, and, detail views, and every visualization for individual clustering results.

Parameter information view

XCluSim provides an overview of parameters for all clustering results in the parameter information view (Figure 44.1a and Figure 44.2a). This view is vertically divided into subsections, each of which corresponds to an individual clustering algorithm (e.g. “K-means clustering”). Inside each subsection, there are multiple bar charts arranged in a matrix layout. Each bar chart shows the number of clustering results generated by the corresponding algorithm with the corresponding parameter setting. A numerical text label is placed on top of the bar in each bar chart to show the number of clustering results. For example, in Figure 44.1, the parameter information view is divided into more than four subsections (some subsections are hidden under the scroll view) since

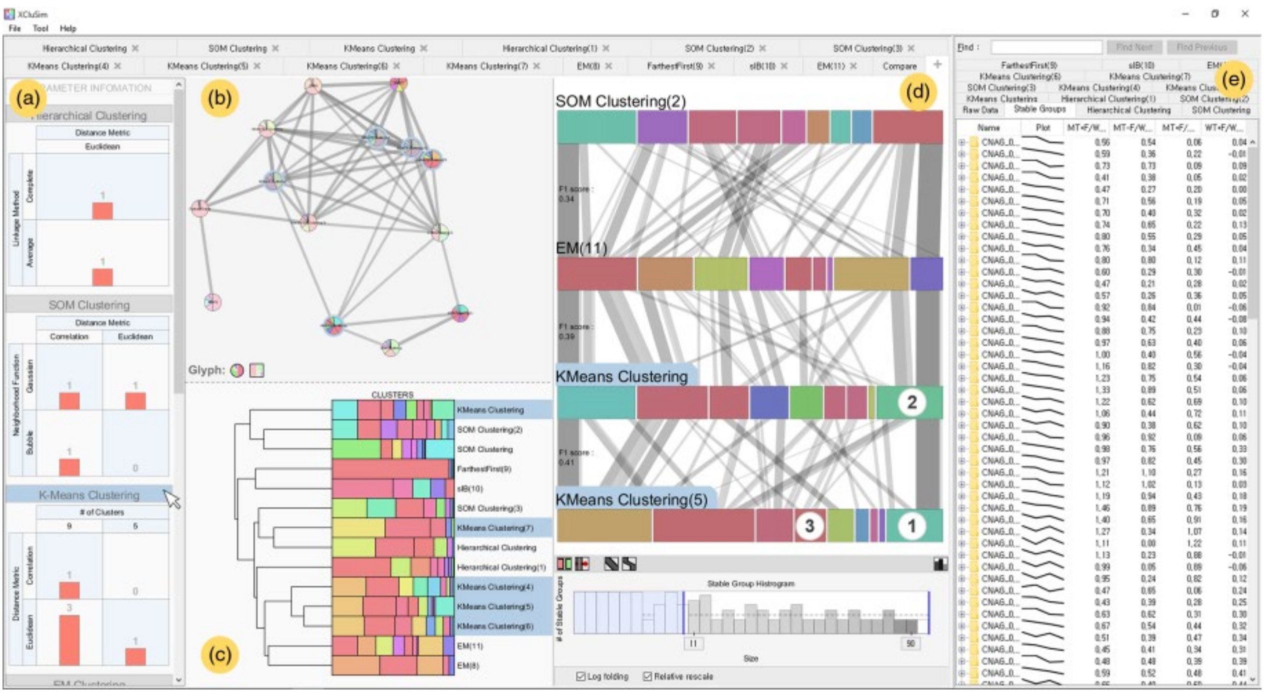


Figure 44.1 Visualization techniques for comparing multiple clustering results in *XCluSim*. There are three types of overviews: (a) parameter information view, (b) force-directed layout overview, and (c) *dendrogram overview*. They enable users to simultaneously compare multiple clustering results in a scalable way. When some clustering results are selected in the overviews, they are added to (d) the *enhanced parallel sets view* for more in-depth comparison tasks. Users can access the detailed information of the selected clustering results with each result in each tab of (e) the *tabular list view*.

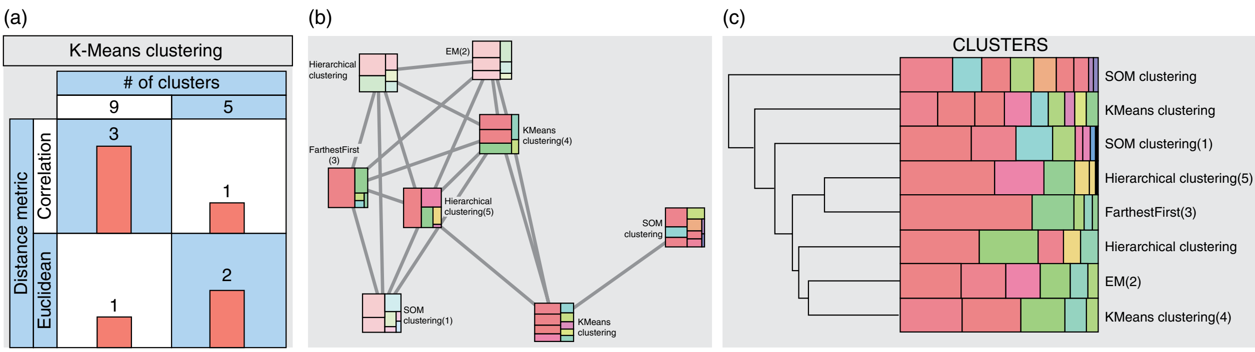


Figure 44.2 Three overviews supported in *XCluSim*. (a) The parameter information view provides the parameter settings used for the clustering results produced. The table in the parameter information view is for a clustering algorithm, and it shows a bar in each cell to represent the number of clustering results using the corresponding parameter setting. (b) The force-directed layout overview intuitively shows similarity among multiple clustering results with the distance between nodes representing similarity. (c) The *dendrogram overview* shows similarities between clustering results in a familiar dendrogram layout with a clustering result visualized at a terminal node.

a user made clustering results using algorithms such as HAC, self-organizing map (SOM) clustering, K-means clustering, and expectation-maximization (EM) clustering. As shown in Figure 44.1, the bar in the left bottom cell of K-means clustering is taller than any bars shown in any clustering algorithms, indicating that the K-means the clustering algorithm with a *distance measure* of Euclidean distance and with 9 as *the number of clusters* is the one mostly used (Figure 44.1). We note, here, that bioinformaticians often run a clustering algorithm multiple times even with the same parameter setting when the algorithm (e.g. K-means) works nondeterministically. For more details on clustering parameters, the user can also look into the visualization of individual clustering results.

To help users determine which results to select for detailed analysis, *XCluSim* provides scalable similarity overviews both at the cluster level and at the clustering result level using a force-directed layout (FDL) and a dendrogram view. In the next two sections, we present details of these two overviews.

Force-directed layout (FDL) overview

In the *FDL overview*, overall similarity relations among multiple clustering results are visualized in a force-directed layout, where more similar results are placed closer together and connected with thicker edges (Figure 44.1b and Figure 44.2b). The similarity metric for calculating distances between nodes is F-measure (Van Rijsbergen, 1974), which is the harmonic mean of the precision and recall measure. Each of the precision and recall measures for the two clustering results is calculated by dividing the number of agreed pairs of items by the number of all pairs of items belonging to a clustering result. An agreed pair refers to two items that "agree" to be clustered together in both clustering results.

Since the *FDL overview* uses physical distance to encode similarity visually between clusters, it has a perceptual advantage in revealing similarity relations among them. In addition, a pie chart is embedded in each node to enable users to visually estimate the number of clusters and their sizes (Figure 44.1b). However, it is challenging to compare angles when a pie chart is too small. To mitigate this problem in the pie chart glyph, users can alternatively visualize each node with a treemap glyph (Bruls, Huizing, & Van Wijk, 2000) (Figure 44.2b). Since the global color encoding scheme also helps users to grasp similarities among clusters, users can estimate which clusters remain stable across different clustering results. For the scalability of the *FDL overview*, nodes become smaller as more results are added to the view. Moreover, an edge between two clusters is displayed only when the similarity between the clusters exceeds a predetermined similarity threshold.

Dendrogram overview

The overall similarity relations are also visualized in the *dendrogram overview* (Figure 44.1c and Figure 44.2c) after running an HAC with all clustering results (i.e. each row or node represents a result). As in the *FDL overview*, we use the F-measure as the distance measure between a pair of results. However, the visual representation and its purpose are different from the *FDL overview*. While the *FDL overview* intuitively shows similarities using physical distance, the *dendrogram overview* uses a more familiar clustering visualization component (i.e. a dendrogram) to represent similarities between clustering results. Moreover, the *dendrogram overview* is more space efficient so that users can see clustering results and cluster distributions more clearly without occlusion.

When users identify clustering results of their interests in the overview of all results, they want to select them and perform more in-depth comparison with them. In the next two subsections, we introduce visualizations for comparing the selected clustering results: the *enhanced parallel sets view* and the *tabular list view*. When a user selects a result either in the *FDL* or *dendrogram overview*s, the selected result is added to the *enhanced parallel sets view* for more in-depth comparison. The *tabular list view*, located on the rightmost side of *XCluSim*, enables users to access detailed information of the selected clustering results with each result in a separate tab.

Enhanced parallel sets view

To visualize the concordance and discordance of multiple clustering results in more detail, we utilized parallel sets (Bendix, Kosara, & Hauser, 2005). We enhanced the parallel sets for effective clustering result comparison by designing more appropriate interactions and revealing more relevant information, i.e., *stable group* (explained in detail later in this section). In the *parallel sets view* (Figure 44.1d and Figure 44.3), each horizontal row of stacked bars represents a clustering result. A tiny gap is placed between each bar to assist users to correctly perceive a single cluster as adjacent bars can occasionally have similar colors when the *Tree Colors* scheme is used. Rows are arranged in such a way that the distance between adjacent rows encodes the dissimilarity between the corresponding clustering results. Each horizontal bar in a row represents a cluster in the corresponding result. We define a *stable group* of items as a set of items that are clustered together through all selected clustering results. A *stable group* is represented as a ribbonlike band across all rows. Since the parallel sets view only enables comparisons based on a linear ordering of results, users can interactively switch any two rows by dragging one over the other. When the vertical order of the rows is changed, all rows are replaced accordingly to reflect the similarity between new adjacent clustering results.

The aggregated band representation for links connecting items in a *stable group* significantly reduces visual clutter compared to the use of a single line representation to connect individual items. The width of a band is an important visual cue that encodes important information about a *stable group* (i.e. its size) in *XCluSim*. Users can easily recognize the largest groups of items that are clustered together across multiple clustering results as they spot thick bands. Moreover, users can estimate the stability of a cluster visually by looking at the width of each *stable group* in it. For example, since the average width of *stable group*s in ① is bigger than ② in Figure 44.3a, a user can infer that ① is a more stable cluster than ②. Cluster similarity based on the color coding of bars (i.e. clusters) helps to facilitate the comparison of multiple clustering results.

However, the aggregation method could still suffer from clutter due to band crossings. We applied a rearrangement algorithm (Zhou et al., 2009) to address this issue. To provide more flexible user interaction depending on a user's need, we divided the algorithm into two rearrangement features: rearranging clusters (i.e. bar rearrangement) and rearranging their members (i.e., band rearrangement). These features can be evoked by pressing on the button at the bottom of the *enhanced parallel sets view* (Figure 44.1d). When a user uses any of these two features, smooth

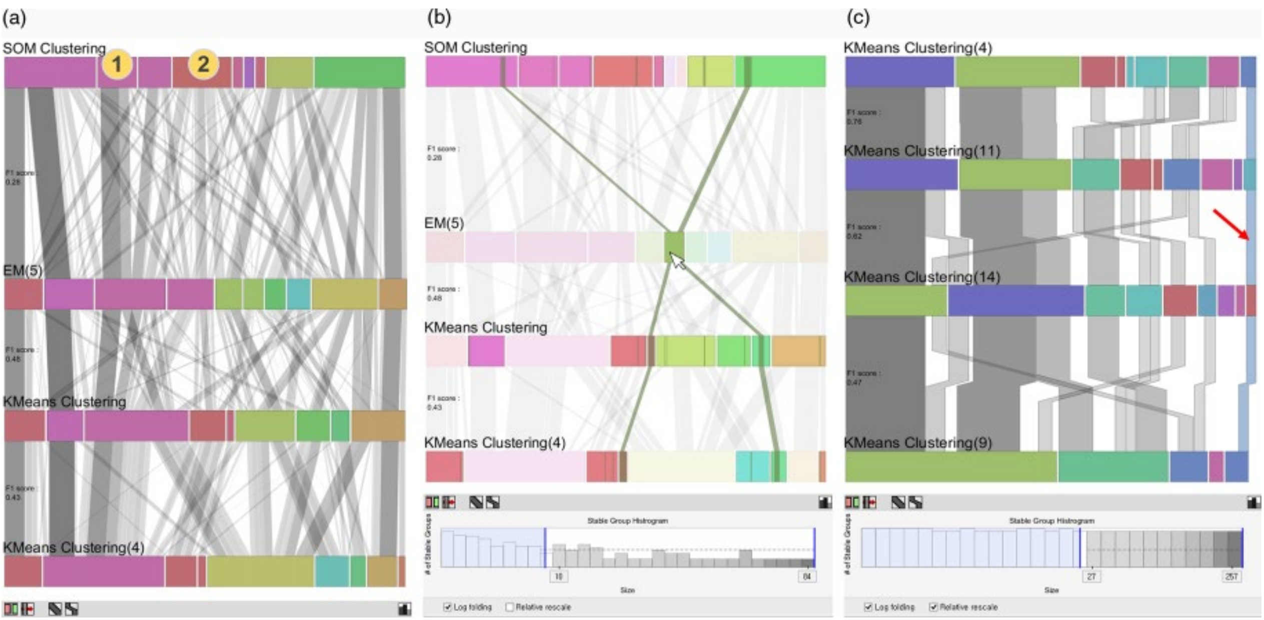


Figure 44.3 The *enhanced parallel sets view* with various user interactions for in-depth comparison. (a) The *parallel sets view* provides rearranging algorithms that minimize line crossings. (b) When users hover a mouse pointer over the node of a cluster, the *stable groups* contained in it are highlighted while other *stable groups* fade out to reveal flows more clearly. By using a filtering feature on the *stable group* histogram at the bottom of the *parallel sets view*, users can hide less interesting bands. (c) Moreover, by using common angle plot (Hofmann, Vendettuoli, 2013), users can compare the sizes of different bands more accurately.

animated transition is supported to reduce the cognitive burden that accompanies users' attempts to trace the movement of bands or bars.

XCluSim provides more user interactions to overcome the cluttering problem. First, users can alleviate the visual clutter in the region of interest by rearranging the bars in a row. This involves dragging them horizontally. After manually rearranging bars (i.e. clusters), users can employ the band rearrangement feature to reduce the visual clutter of bands across multiple rows due to the current manual arrangement of bars in the row. Secondly, there is a band filtering feature similar to that in Lex, Schulz, Streit, Partl, and Schmalstieg (2011). The *stable group* histogram at the bottom of Figure 44.3c shows the distribution of bands by size. There are two blue filtering bars on both sides. Users can filter out bands that are too small or too big from the *parallel sets view* by adjusting the position of the filtering bars. Finally, when the mouse pointer hovers over a cluster, it highlights the bands, allowing the clusters to show their flows across other clustering results clearly (Figure 44.3b). This can be helpful when a user is especially interested in *stable groups* that belong to a specific cluster.

The perception of a stable group's size could be distorted by a line-width illusion (Hofmann & Vendettuoli, 2013). Such an illusion causes humans to perceive line width incorrectly at slanted angles. This distortion may disrupt the task of band size comparison. In order to prevent it, we adopt the common angle plot (Hofmann & Vendettuoli, 2013) idea (Figure 44.3c). By comparing the straight, vertical parts of bands, users can compare the sizes of the *stable groups* more accurately. However, since the common angle plot represents a single line as three connected straight lines, it may generate more clutter and occlusions. Thus, it is better to use this feature when only a small number of bands are displayed in the *parallel sets view*.

Tabular list view

Users can access detailed information concerning the selected clustering results with each result in a separate tab in the *tabular list view* (Figure 44.4). The tabular view provides detailed information in two different modes: the group-by mode and the

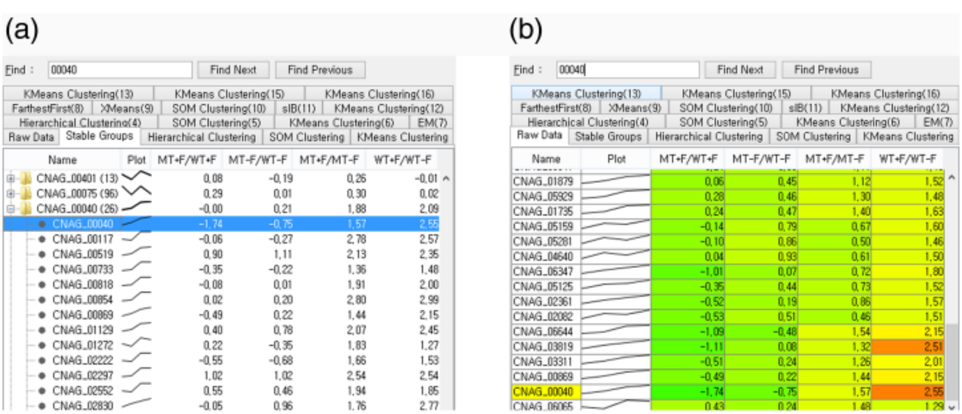


Figure 44.4 The *tabular list view* enables users to access numerical details. (a) Users can see detailed information for each item grouped by cluster or *stable group*. (b) Users also can see raw data in a heatmap form. When a user wants to access an item or a group directly, he / she can use the search box provided on top of the *tabular list view*.

heatmap mode. In the group-by mode, users can see the data grouped by *stable groups* or by clusters. A group is represented by a representative item in a single row with the number of group members between parentheses. Moreover, there is a line graph glyph in each row to show the overall average pattern of the corresponding group. In the heatmap mode, the *tabular list view* shows numerical details with each cell color coded according to its value. There is a text search field on top of the *tabular list view* so that users can access specific items directly. A user can export a selected subset of data (e.g. a specific *stable group*) as a CSV text file for further analysis.

XCluSim provides brushing and linking among all visualization components. Thus, the *tabular list view* is coordinated with all visualization components in *XCluSim*. Thus, whenever a user selects a group of items in any visualization, they are highlighted in the *tabular list view* to help the user access detailed information about them. In addition, when the mouse pointer hovers over an item in a component, it highlights the item in white-blue color, and all related items on the other components are also highlighted. This could lead to additional meaningful insights. For example, hovering a mouse pointer over the title of a specific algorithm in the parameter information view results in the highlighting of all related clustering results in overviews and detail views (Figure 44.1). As a consequence, users are able to understand that K-means clustering can produce totally different clustering results depending on the clustering parameters chosen (e.g. compare “K-means clustering” to “K-means clustering(4)” in the *dendrogram overview* in Figure 44.1).

Interactive data manipulation

Simple file formats such as comma separated values (CSV) and tab-delimited text are used for *XCluSim*. *XCluSim* enables researchers to interactively manipulate the input dataset when loading it, prior to clustering it (Figure 44.5). Users can generate a ratio value by selecting two columns from the original dataset. *XCluSim* provides filters such as a range filter and RPKM threshold adjustment. It also provides features for calculating fold changes.

Clustering algorithms supported in XCluSim

To make *XCluSim* a more general visual analytics tool for comparing clustering results, we try to provide a wide variety of clustering algorithms. First, we implement frequently used clustering algorithms in *XCluSim*. These include hierarchical agglomerative clustering (Eisen et al., 1998), SOM clustering (Kohonen, 1990), K-means clustering, and OPTICS clustering (Ankerst, Breunig, Kriegel, & Sander, 1999). Moreover, all clustering algorithms from Weka (Hall et al., 2009) are also available in *XCluSim*. Users can also import any clustering results made by any other clustering algorithms that are not available in *XCluSim*.

Visualization technique for hierarchical clustering

We visualized HAC results with the combination of a dendrogram and heatmap visualization (Figure 44.6a), where users could interactively compress / expand, flip, and swap subtrees. The batch compression of subtrees using the minimum similarity bar

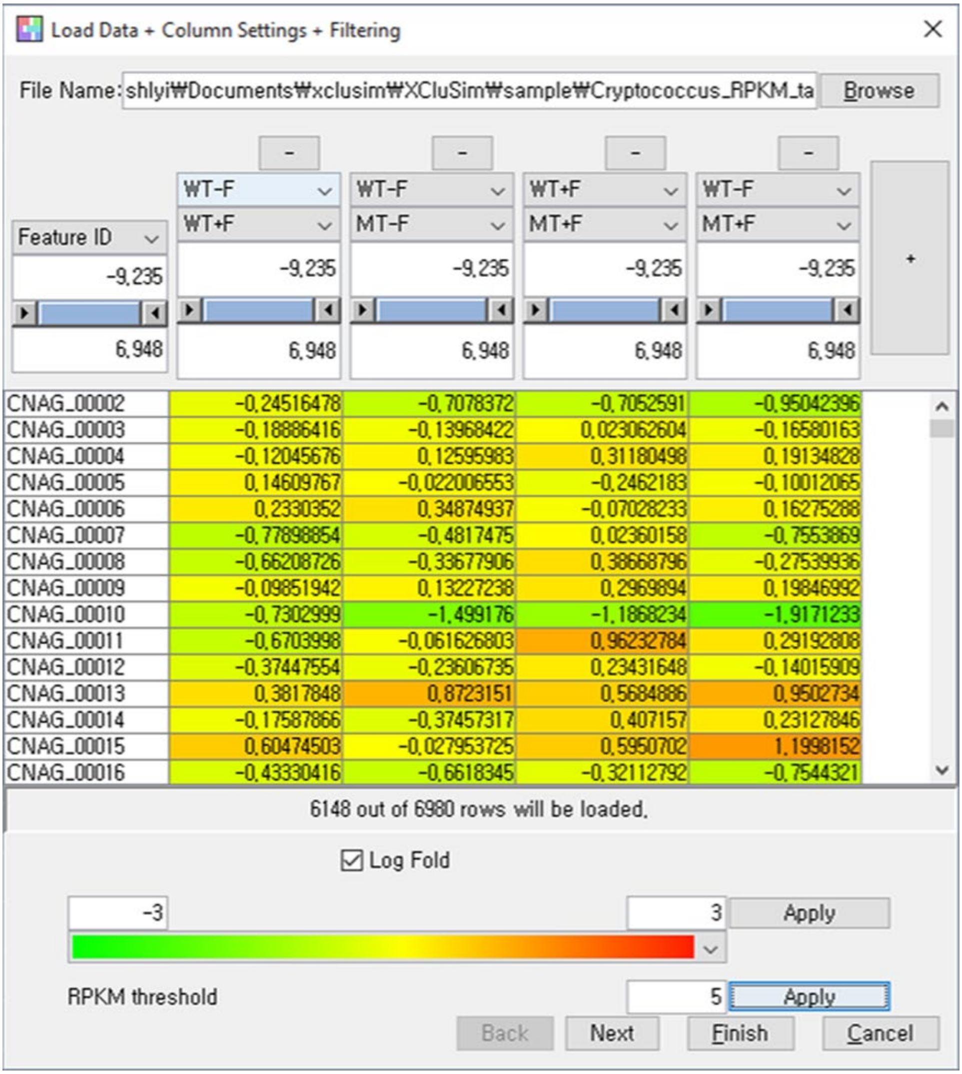


Figure 44.5 Interactive manipulation of input data supported by *XCluSim*: derive a new column (ratio, fold change), change color mapping, filter items using a range filter and RPKM adjustment.

(Seo & Shneiderman, 2002) is also possible. By adjusting the position of the similarity bar, users can dynamically determine the clusters. There is a compact bird's-eye overview using heatmap (Lex, Streit, Kruijff, & Schmalstieg, 2010) in the leftmost part, which is tightly coupled with the dendrogram. By dragging a black-bordered rectangle that represents the current viewport (see the black rectangle in the top left of Figure 44.6a) in the heatmap overview, users can efficiently navigate through the dendrogram + heatmap view.

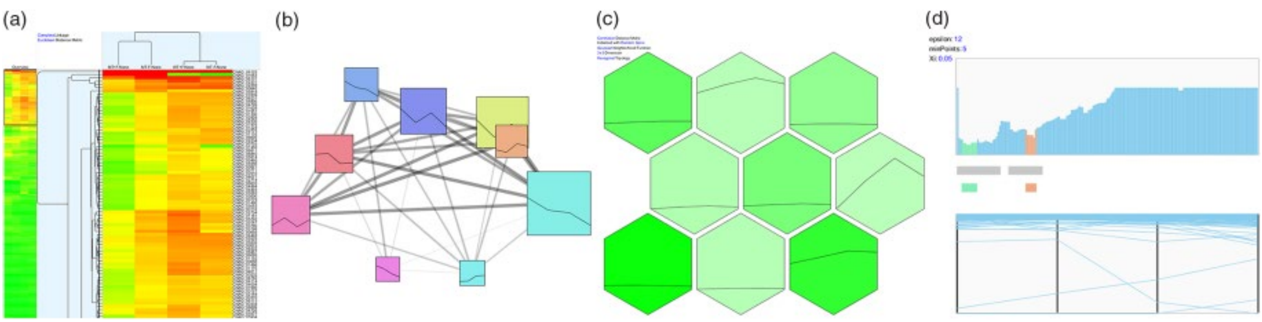


Figure 44.6 Visualization techniques for individual clustering results in *XCluSim*. (a) Dendrogram+heatmap visualization for hierarchical agglomerative clustering results. (b) Force directed layout for every partitional clustering result and imported clustering results. (c) Common hive-shaped visualization for SOM clustering results. (d) Reachability plot together with parallel coordinate plot for OPTICS clustering results.

Visualization technique for partitional method

Partitional clustering results other than SOM clustering (e.g. K-means clustering, EM clustering, farthest first clustering, etc.), and all imported results are visualized in a force-directed layout (Figure 44.6b), where each cluster is represented as a rectangle whose size is proportional to the cluster size. The force between nodes is determined by the similarity between members of each cluster so that similar clusters are closely positioned and have thicker links between them. To show an overview of a cluster, *XCluSim* also visualizes the average pattern of all members of the cluster in a line chart, which is shown as a glyph in the cluster's node. *XCluSim* also supports semantic zooming to enable users to explore clusters in more detail. When a cluster is zoomed into, more details of its members are dynamically visualized in a parallel coordinate plot.

SOM clustering results are visualized using the typical hive-shaped visualization (Figure 44.6c), where each hexagonal cell represents a cluster. In *XCluSim*, the background intensity of each cell represents the size of the corresponding cluster. As a visual summary of each cluster, *XCluSim* presents the average pattern of the cluster members in a line chart within each hexagonal cell. *XCluSim* also supports semantic zooming. Users can zoom into a cluster by double clicking on the corresponding cell and look at the details of their members in a parallel coordinate plot in the same way they would in a force-directed layout.

Visualization Technique for Density-Based Method

Density-based clustering algorithms calculate a kind of density-related information for each item during the clustering process. For example, OPTICS (Ankerst, Breunig, Kriegel, & Sander, 1999) calculates the reachability distance for each item. We believe that users can more intuitively understand a density-based clustering result when the density-related information is revealed. Therefore, a bar-chart-like visualization, with each item arranged on the horizontal axis and the density-related information on the vertical axis, can effectively visualize density-based clustering results. The conventional reachability plot for OPTICS is a typical example. In *XCluSim*, we enhance the plot for better cluster identification and for improved examination of details (Figure 44.6d). To show the position of each cluster clearly, *XCluSim* places a horizontal bar from the start to the end positions of the cluster right below the reachability plot. The parallel coordinate plot at the bottom shows more details of cluster members. These two plots support brushing and linking between the cluster members. For example, when a mouse pointer hovers over a cluster in the reachability plot, the lines for the members of the cluster are highlighted in the parallel coordinate plot.

Implementation

XCluSim was developed using Java Standard Edition 7 (Java SE 7), which enables it to run on any platform with JRE version 1.7 or higher. We used the Piccolo 2D framework to implement visualization components and interactions. Weka's clustering algorithms were integrated into *XCluSim* using Weka SDK 3.6 (Hall et al., 2009).

Case Studies

To evaluate the efficacy of *XCluSim*, we conducted two case studies with our collaborator in a major bioinformatics research laboratory. He is a senior research engineer and has years of experience in genome and transcriptome analyses.

Elucidating the role of ferroxidase in *Cryptococcus neoformans* var. grubii H99 (Case Study 1)

This study was carried out in his laboratory for 80 min. Prestudy and poststudy interviews were conducted for 10 min each. The participant used *XCluSim* for 50 min after a 10 min tutorial. We used a dataset containing normalized expression levels of 6,980 genes belonging to the *Cryptococcus neoformans var. grubii H99* strain. The dataset had been prepared for his previous work (Kim et al., 2012).

His task was to elucidate the role of ferroxidase (cfo1) by knocking it out. He was interested in finding a meaningful set of genes whose expression would be influenced and in identifying the affected pathways. For the task, he tried to see the effect of fluconazole on two different strains: the wild type of *Cryptococcus neoformans var. grubii H99* and the cfo1 mutant of the same strain. In the dataset, each gene has four expression levels: two different strains, each cultured in two conditions (i.e. wild-type strain and cfo1 mutant with and without fluconazole treatment).

When he loaded the data, he made four new data columns of ratio values, including the wild-type strain with fluconazole versus the wild-type strain without fluconazole treatment (WT+F/WT-F) and the cfo1 mutant with fluconazole versus the cfo1 mutant without fluconazole treatment (MT+F/MT-F) (Figure 44.5). Subsequently, he adjusted the RPKM threshold and used log fold changes to filter out less interesting genes for more efficient analysis.

After data preprocessing, *XCluSim* showed the results of three clustering algorithms (i.e. HAC, SOM clustering, K-means clustering) in three independent views. He was most familiar with dendrogram and heatmap visualizations, so he examined the HAC results first. He was interested in genes that were highly expressed with fluconazole treatment. Among them he found the gene named Erg11 (CNAG_00040). He said that this gene was reported to be associated with azole resistance.

Next, he tried to see which genes were stably grouped together across different clustering results. He tried to load as many clustering results as possible to see the differences between them. The parameter information view provided him with a good overview of all clustering results (clustering algorithms and their parameters). He was able to make diverse clustering results without generating any duplicate results.

After generating 15 different clustering results, he selected four diverse results from the *FDL overview* to find out which genes were clustered together with Erg11. However, he recognized that the *stable group*s were excessively thin because of the result named "FarthestFirst(6)." This had to do with the fact that it was the most dissimilar result to other selected clustering results. So he removed that result from the *parallel sets view*. Then he selected a more similar one named "KMeans Clustering(4)" (Figure 44.3a). He subsequently accessed the *stable group* with Erg11 directly, utilizing the search feature in the *tabular list view*. He was able to confirm that 17 other genes belonged to the *stable group*. After validating the members of the *stable group* with an enrichment analysis, he found that most of them (10 out of 18) belonged to the ergosterol biosynthetic pathway.

Once he had selected the *stable group* in the *tabular list view*, he was able to efficiently inspect the flow of the group across different clustering results in the *enhanced parallel sets view* (Figure 44.3b). While he looked into the flow of the *stable group* across all rows (the rightmost highlighted band in Figure 44.3b), he also noticed that the clustering result from "KMeans Clustering(4)" had the tightest cluster, which included the *stable group*. However, there were no more genes outside the *stable group* in the cluster that belonged to the ergosterol pathway.

Then he tried to find the best algorithm and those of its parameters that gave the tightest cluster containing genes belonging to the ergosterol pathway. Since "KMeans Clustering(4)" had previously been the best clustering result among the selected results, he ran K-means clustering algorithms with different parameters to arrive at similar results. He then inserted three of the most similar results in the parallel sets view (Figure 44.3c). Again, he highlighted a *stable group* with Erg1 (the band indicated with a red arrow in Figure 44.3c). By checking the flow of the *stable group* crossing each result, he recognized that "KMeans Clustering(14)" gave the tightest cluster. This led to the conclusion that K-means clustering with the corresponding parameter configurations (i.e. Euclidean distance as the distance metric and nine as the number of clusters) was the best result for the given dataset among all the results.

Finding a clustering result that clearly represents biological relations (Case Study 2)

A second case study was subsequently carried out with the same participant in his laboratory. The study was conducted for 150 min on a different day. The participant was already familiar with *XCluSim*, so we skipped the tutorial. In the study, he relied on the gene expression profiles of 169 genes in *Escherichia coli*, which used a DNA microarray (Khodursky et al., 2000). In the dataset, each gene contained 19 expression levels in order to investigate the effects of the perturbations on tryptophan metabolism. The expressions were measured under the following conditions: wild type growth with and without tryptophan (five conditions), wild type growth with and without tryptophan starvation (nine conditions), and the growth of wild type and a *trp* repressor mutant (five conditions).

Through the case study, the participant wanted to find a clustering result that clearly reflected biological relations in tryptophan metabolism. In the original paper (Khodursky et al., 2000), the authors used HAC to cluster the 169 gene expression profiles measured in the 19 conditions. It was indicated in the paper that genes showing similar expression responses did not necessarily fall into the same cluster. One example included the genes associated with aromatic amino acid metabolism.

He first wanted to see if the optimal algorithm and its parameters in the previous case study would work for another dataset. To determine this, he produced 11 clustering results in *XCluSim*, including the result produced using previous optimal settings: K-means clustering with Euclidean distance as the *distance metric* and 9 as the *number of clusters*. He validated each cluster in the result ("KMeans Clustering(6)" in Figure 44.7a) through an enrichment analysis using the DAVID

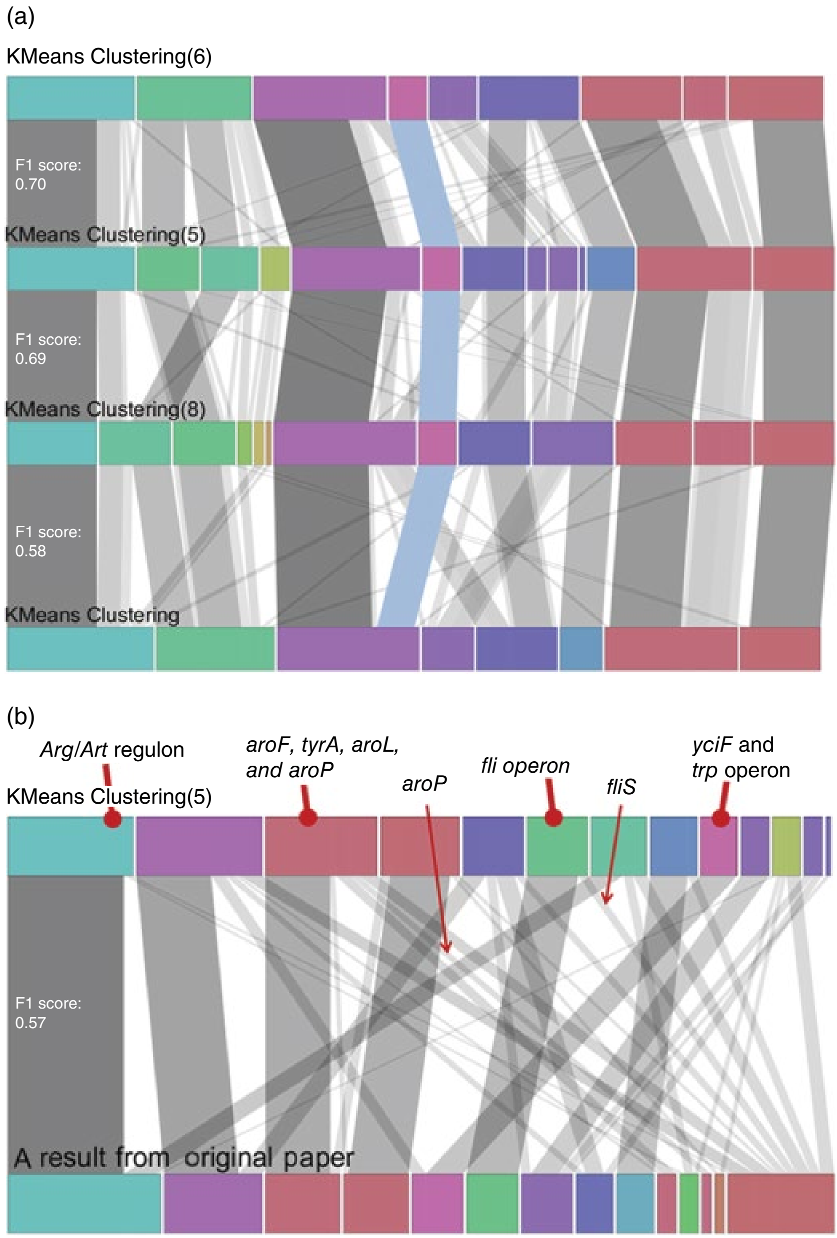


Figure 44.7 Results of the second case study are visualized in the *enhanced parallel sets view*. (a) The highlighted *stable group* contained the trp operon with yciF. (b) Visual comparison of two results: the best clustering result ("KMeans Clustering(5)") derived from the case study and a result ("A Result from Original Paper") presented in the original research paper (Khodursky et al., 2000).

website (http://david.abcc.ncifcrf.gov/). After validating each cluster, he concluded that most of the clusters were grouped well in the sense that they represented biological relations in pathways. However, he recognized two problems in the result. First, a cluster that had both *Arg* and *Art* regulons also contained a gene named *tnaA* that was considered to be noise. This was because *tnaA* showed a different expression pattern and was not highly related to other cluster members in biological terms. Secondly, one gene from the *fli* operon, *fliS*, fell into a different cluster from the other genes in the same operon while they had homogeneous expression patterns.

By utilizing visualizations in *XCluSim*, he wanted to find the clustering result that properly represented biological relations as "KMeans Clustering(6)" while the two problems were revisited. For this intended task, he selected all the similar results from the *FDL overview*: "KMeans Clustering(5)," "KMeans Clustering(8)," and "KMeans Clustering." Then he accessed the *stable group*s that contained *tnaA* and the *Arg/Art* regulon. He easily recognized that genes in both the *Arg* and *Art* regulons fell into the same *stable group* while *tnaA* was not stably clustered with them. The results, which separately clustered *tnaA* from the *Arg* and *Art* regulons, were "KMeans Clustering(5)" and "KMeans Clustering(8)." Similarly, by checking the flow of *stable group*s in each horizontal row, he easily recognized that two clustering results that used the correlation coefficient as a distance metric clustered two *stable group*s together: one with the *fli* operon and the other with *fliS*. The two results were "KMeans Clustering(5)" and "KMeans Clustering." As a consequence, "KMeans Clustering(5)," using the correlation coefficient as the distance metric and 13 as the number of clusters was the most satisfying result for the dataset.

Additionally, our participant gained insight by seeing a *stable group* in *XCluSim*. Genes in the *trp* operon (i.e. *trpE*, *trpD*, *trpC*, *trpB*, and *trpA*) were stably clustered together with *yciF* through the four different results (see the highlighted *stable group* in Figure 44.7a). As *yciF* was assigned to a putative function, he said that the gene might be closely related to tryptophan synthase as a *trp* operon.

After he found the best result, he compared it with a clustering result provided in the original work (Khodursky et al., 2000) to see if his result better represented biological relations (Figure 44.7b). The clustering result presented in the paper had been prepared prior to the study and was imported to *XCluSim* for visual comparisons. After comparing two results, he found that some of the genes involved in aromatic amino acid metabolism, *aroF*, *tyrA*, *aroL*, and *aroP*, were clustered together in our best result while only three of them fell into the same cluster in their original result. Moreover, their result did not cluster *fliS* with the other *fli* operon. These results suggested that the authors of the original work (Khodursky et al., 2000) could have generated more biologically meaningful results if they had used *XCluSim* in the first place.

Discussion

During the case studies, we received positive subjective feedback on *XCluSim* from the participant. He especially liked the ability to identify *stable group*s across multiple clustering results. Moreover, he was satisfied that he could select and run diverse clustering algorithms and interactively compare them by adding/removing a clustering result to / from the enhanced *parallel sets view*. He could quickly shift his attention to a more interesting set of results for more in-depth comparison. However, he also pointed out the limitations of *XCluSim*. As filtering sets of items was only available at the data manipulation step, he said it would be helpful to allow users to interactively filter raw data in the visualization components as well.

We color-coded each cluster consistently across the whole system using the *Tree Colors* scheme after building a hierarchical structure of all clusters from multiple clustering results. With the help of this color coding, overviews became even more useful in *XCluSim*. While the color encoding was applied for a specific purpose in this work (i.e. for the visualization of clusters), we think it can also be applied to parallel sets applications in a more general and scalable way. For example, instead of distinguishing only a small number of categories while visualizing a categorical dataset, it might be possible to distinguish many more nodes in the parallel sets once a hierarchical structure of the nodes has been built in a similar manner to the one we employed in *XCluSim*.

Future Work

At present, when a clustering algorithm does not assign all items to clusters, all unclustered items are treated as a single cluster in *XCluSim*. OPTICS and DBSCAN clustering algorithms can give rise to results of this kind. *XCluSim* treats unclustered items as a group of less interesting items, as if it were a special cluster. Otherwise, it could make a huge number of *stable group*s because each unclustered item will become a single *stable group*. This would make it hard for users to gain insight from visualizations. In the future, we plan to improve *XCluSim* to resolve this problem. For example, we can represent these kinds of groups with different textures in the *parallel sets view* to distinguish them from other normal clusters.

In this research, we concentrated mostly on supporting comparison tasks based on the concordance / discordance of multiple clustering results. However, since bioinformaticians' cluster analysis is highly integrated with the validation stage, it would also be valuable to provide a visual representation of cluster validity measures (e.g. internal cluster validity indices). For example, the gray scale intensity of each band (i.e. *stable group*) in the *parallel sets view*, which currently represents the size of a *stable group*, can be used to represent its internal validity measures. In such a case, *stable group* provided by *XCluSim* will become more reliable information.

Conclusion

In this research we presented *XCluSim*, a visual analytics tool that enables users to compare multiple clustering results. *XCluSim* provides three different overviews to help users grasp their overall similarity relationships in a more scalable and flexible way. Moreover, the enhanced *parallel sets view* enables users to detect differences among select clustering results even more clearly by using improved user interactions. We conducted case studies to evaluate the usefulness of *XCluSim*, and the participants gave positive feedback.

References

Andrienko, G., Andrienko, N., Rinzivillo, S., Nanni, M., Pedreschi, D., & Giannotti, F. (2009). Interactive visual clustering of large collections of trajectories. In *IEEE Symposium on Visual Analytics Science and Technology, 2009 (VAST 2009)* (pp. 3–10). doi:10.1109/VAST.2009.5332584

Ankerst, M., Breunig, M. M., Kriegel, H. P., & Sander, J. (1999). OPTICS: Ordering points to identify the clustering structure. In *Proceedings of the ACM SIGMOD International Conference on Management of Data* (pp. 49–60). doi:10.1145/304182.304187

Bendix, F., Kosara, R., & Hauser, H. (2005). Parallel sets: Visual analysis of categorical data. In *IEEE Symposium on Information Visualization, 2005 (INFOVIS 2005)* (pp. 133–140). doi:10.1109/INFVIS.2005.1532139

Bruls, M., Huizing, K., & Van Wijk, J. J. (2000). *Squarified treemaps* (pp. 33–42). Vienna: Springer.

Collins, C., Penn, G., & Carpendale, S. (2009). Bubble sets: Revealing set relations with isocontours over existing visualizations. *IEEE Transactions on Visualization and Computer Graphics*, *15*(6), 1009–1016.

Ding, H., Wang, C., Huang, K., & Machiraju, R. (2014). iGPSe: A visual analytic system for integrative genomic based cancer patient stratification. *BMC Bioinformatics*, *15*(1), 203.

Eisen, M. B., Spellman, P. T., Brown, P. O., & Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences*, *95*(25), 14863–14868.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD Explorations Newsletter*, *11*(1), 10–18.

Havre, S. L., Shah, A., Posse, C., & Webb-Robertson, B. J. (2006). Diverse information integration and visualization. In *Proceedings of the SPIE Conference on Visualization and Data Analysis* (pp. 60600M). doi:10.1117/12.643492

Hofmann, H., & Vendettuoli, M. (2013). Common angle plots as perception-true visualizations of categorical associations. *IEEE Transactions on Visualization and Computer Graphics*, *19*(12), 2297–2305.

Hossain, M. S., Ojili, P. K. R., Grimm, C., Müller, R., Watson, L. T., & Ramakrishnan, N. (2012). Scatter/gather clustering: Flexibly incorporating user feedback to steer clustering results. *IEEE Transactions on Visualization and Computer Graphics*, *18*(12), 2829–2838.

Inselberg, A., & Dimsdale, B. (1990). Parallel coordinates: a tool for visualizing multi-dimensional geometry. In *IEEE Symposium on Information Visualization, 1990 (INFOVIS 1990)* (pp. 361–378). Los Alamitos, CA: IEEE.

Kandogan, E. (2012). Just-in-time annotation of clusters, outliers, and trends in point-based data visualizations. In *2012 IEEE Conference on Visual Analytics Science and Technology (VAST)* (pp. 73–82). doi:10.1109/VAST.2012.6400487

Khodursky, A. B., Peter, B. J., Cozzarelli, N. R., Botstein, D., Brown, P. O., & Yanofsky, C. (2000). DNA microarray analysis of gene expression in response to physiological and genetic changes that affect tryptophan metabolism in *Escherichia coli*. *Proceedings of the National Academy of Sciences*, *97*(22), 12170–12175.

Kim, J., Cho, Y. J., Do, E., Choi, J., Hu, G., Cadieux, B.,…Jung, W. H. (2012). A defect in iron uptake enhances the susceptibility of *Cryptococcus neoformans* to azole antifungal drugs. *Fungal Genetics and Biology*, *49*(11), 955–966.

Kohonen, T. (1990). The self-organizing map. *Proceedings of the IEEE*, *78*(9), 1464–1480.

Kothur, P., Sips, M., Dobslaw, H., & Dransch, D. (2014). Visual analytics for comparison of ocean model output with reference data: Detecting and analyzing geophysical processes using clustering ensembles. *IEEE Transactions on Visualization and Computer Graphics*, *20*(12), 1893–1902.

Lex, A., Schulz, H. J., Streit, M., Partl, C., & Schmalstieg, D. (2011). VisBricks: Multiform visualization of large, inhomogeneous data. *IEEE Transactions on Visualization and Computer Graphics*, *17*(12), 2291–2300.

Lex, A., Streit, M., Kruijff, E., & Schmalstieg, D. (2010). Caleydo: Design and evaluation of a visual analysis framework for gene expression data in its biological context. In *Proceedings of the IEEE Symposium on Pacific Visualization* (pp. 57–64). doi:10.1109/PACIFICVIS.2010.5429609

Lex, A., Streit, M., Partl, C., Kashofer, K., & Schmalstieg, D. (2010). Comparative analysis of multidimensional, quantitative data. *IEEE Transactions on Visualization and Computer Graphics*, *16*(6), 1027–1035.

Lex, A., Streit, M., Schulz, H. J., Partl, C., Schmalstieg, D., Park, P. J., & Gehlenborg, N. (2012). StratomeX: Visual Analysis of Large-Scale Heterogeneous Genomics Data for Cancer Subtype Characterization. *Computer Graphics Forum*, *31*(3), 1175–1184.

L'Yi, S., Ko, B., Shin, D., Cho, Y. J., Lee, J., Kim, B., & Seo, J. (2015). XCluSim: A visual analytics tool for interactively comparing multiple clustering results of bioinformatics data. *BMC Bioinformatics*, *16*(suppl 11), S5.

Pilhöfer, A., Gribov, A., & Unwin, A. (2012). Comparing clusterings using Bertin's idea. *IEEE Transactions on Visualization and Computer Graphics*, *18*(12), 2506–2515.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65.

Seo, J., & Shneiderman, B. (2002). Interactively exploring hierarchical clustering results. *Computer*, *35*(7), 80–86.

Sharko, J., Grinstein, G. G., Marx, K., Zhou, J., Cheng, C. H., Odelberg, S., & Simon, H. G. (2007). Heat map visualizations allow comparison of multiple clustering results and evaluation of dataset quality: Application to microarray data. In *Proceedings of the IEEE 11th International Conference Information Visualization* (pp. 521–526). doi:10.1109/IV.2007.61

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages* (pp. 336–343). doi:10.1109/VL.1996.545307

Tennekes, M., & de Jonge, E. (2014). Tree colors: color schemes for tree-structured data. *IEEE Transactions on Visualization and Computer Graphics*, *20*(12), 2072–2081.

Van Rijsbergen, C. J. (1974). Foundation of evaluation. *Journal of Documentation*, *30*(4), 365–373.

Zhou, J., Konecni, S., & Grinstein, G. (2009). Visually comparing multiple partitions of data with applications to clustering. In *Proceedings of the SPIE Conference on Visualization and Data Analysis* (p. 72430J). doi:10.1117/12.810093

Index

Entries from page numbers 1 to 514 are from Volume 1 and page numbers 515 to 966 are from Volume 2
Page numbers in *italics* refer to Figures
Page numbers in **bold** refer to Tables

The Wiley Handbook of Human Computer Interaction, Volume 2, First Edition.
Edited by Kent L. Norman and Jurek Kirakowski.