LNCS 6949

Pedro Campos Nicholas Graham
Joaquim Jorge Nuno Nunes
Philippe Palanque Marco Winckler (Eds.)

Human-Computer Interaction – INTERACT 2011

13th IFIP TC 13 International Conference
Lisbon, Portugal, September 2011
Proceedings, Part IV

Part IV

ifip

Springer

Lecture Notes in Computer Science 6949

Pedro Campos Nicholas Graham
Joaquim Jorge Nuno Nunes
Philippe Palanque Marco Winckler (Eds.)

Human-Computer Interaction – INTERACT 2011

13th IFIP TC 13 International Conference
Lisbon, Portugal, September 5-9, 2011
Proceedings, Part IV

Springer

Volume Editors

Pedro Campos
Nuno Nunes
University of Madeira
9020-105, Funchal, Portugal
E-mail: {pcampos, njn}@uma.pt

Nicholas Graham
Queen's University
Kingston, ON K7L 3N6, Canada
E-mail: graham@equis.cs.queensu.ca

Joaquim Jorge
Instituto Superior Técnico
1049-001 Lisbon, Portugal
E-mail: jaj@inesc.pt

Philippe Palanque
Marco Winckler
University Paul Sabatier
31062 Toulouse Cedex 9, France
E-mail: {palanque, winckler}@irit.fr

ISSN 0302-9743 e-ISSN 1611-3349
ISBN 978-3-642-23767-6 e-ISBN 978-3-642-23768-3
DOI 10.1007/978-3-642-23768-3
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2011935338

CR Subject Classification (1998): H.5.2, H.5.3, H.3-5, I.2.10, D.2, C.2, K.3-4

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Foreword

Advances in interactivity, computing power, mobile devices, large displays and ubiquitous computing offer an ever-increasing potential for empowering users. This can happen within their working environment, in their leisure time or even when extending their social skills. While such empowerment could be seen as a way of connecting people in their workspace, home or on the move, it could also generate gaps requiring larger effort and resources to fruitfully integrate disparate and heterogeneous computing systems.

The conference theme of INTERACT 2011 was "building bridges" as we believe human–computer interaction (HCI) is one the research domains more likely to significantly contribute to bridging such gaps. This theme thus recognizes the interdisciplinary and intercultural spirit that lies at the core of HCI research. The conference had the objective of attracting research that bridges disciplines, cultures and societies. Within the broad umbrella of HCI, we were in particular seeking high-quality contributions opening new and emerging HCI disciplines, bridging cultural differences, and tackling important social problems. Thus, INTERACT 2011 provided a forum for practitioners and researchers to discuss all aspects of HCI, including these challenges. The scientific contributions gathered in these proceedings clearly demonstrate the huge potential of that research domain to improving both user experience and performance of people interacting with computing devices. The conference also is as much about building bridges on the human side (between disciplines, cultures and society) as on the computing realm.

INTERACT 2011 was the 13th conference of the series, taking place 27 years after the first INTERACT held in early September 1984 in London, UK. Since INTERACT 1990 the conferences have taken place under the aegis of the UNESCO International Federation for Information Processing (IFIP) Technical Committee 13. This committee aims at developing the science and technology of the interaction between humans and computing devices through different Working Groups and Special Interests Groups, all of which, together with their officers, are listed within these proceedings.

INTERACT 2011 was the first conference of its series to be organized in cooperation with ACM SIGCHI, the Special Interest Group on Computer–Human Interaction of the Association for Computing Machinery. We believe that this cooperation was very useful in making the event both more attractive and visible to the worldwide scientific community developing research in the field of HCI.

We thank all the authors who chose INTERACT 2011 as the venue to publish their research This was a record year for the conference in terms of submissions in the main technical categories. For the main Technical Program there were a total of 680 submissions, including 402 long and 278 short papers, out of which we accepted 171 (111 long and 60 short submissions), for a combined acceptance rate of less than 25%. Overall, from a total of 741 submissions for all tracks, 290 were accepted, as follows:

- 111 Full Research Papers
- 60 Short Research Papers
- 54 Interactive Poster Papers
- 17 Doctoral Consortium Papers
- 16 Workshops
- 12 Tutorials
- 5 Demonstrations
- 6 Organizational Overviews
- 4 Industrial Papers
- 3 Special Interest Groups
- 2 Panels

Our sincere gratitude goes to the members of our Program Committee (PC), who devoted countless hours to ensure the high quality of the INTERACT Conference. This year, we improved the reviewing process by moving to an associate chair model. With almost 700 submitted papers, it is impossible for the PC Chairs to read every paper. We recruited 103 Associate Chairs (ACs), each of whom handled up to 12 papers. The ACs recruited almost 800 external reviewers, guaranteeing that each paper was reviewed by three to six referees. ACs also provided a meta-review. Internal discussion among all the reviewers preceded the final decision between the PC Chairs and the AC. This herculean effort was only possible due to the diligent work of many people. We would like to thank you all for the effort and apologize for all the bullying required to get the work done on time.

In addition, sincere thanks must be extended to those whose contributions were essential in making it possible for the conference to happen and for these proceedings to be produced. We owe a great debt to the Conference Committees, the members of the International Program Committee and the numerous reviewers who had to review submissions from the various categories. Similarly, the members of the conference Organizing Committee, the staff at INESC-ID, especially Manuela Sado, deserve much appreciation for their tireless help with all aspects of planning and managing the many administrative and organizational issues. We would like to especially thank Tiago Guerreiro for his dedication with the Student Volunteer program, and José Coelho who worked tirelessly to make the online program a reality. Thanks are also due to Alfredo Ferreira for keeping and single-handedly maintaining the website, and to Pedro Campos and Marco Winkler for the superb work done with the conference proceedings. Finally, our thanks go to all the authors who actually did the scientific work and especially to the presenters who took the additional burden of discussing the results with their peers at INTERACT 2011 in Lisbon.

July 2011

Nicholas Graham
Daniel Gonçalves
Joaquim Jorge
Nuno Nunes
Philippe Palanque

IFIP TC13

Established in 1989, the International Federation for Information Processing Technical Committee on Human–Computer Interaction (IFIP TC13) is an international committee comprising 30 national societies and 7 working groups, representing specialists in human factors, ergonomics, cognitive science, computer science, design and related disciplines. INTERACT is its flagship conference, staged biennially in different countries in the world.

IFIP TC13 aims to develop the science and technology of human–computer interaction (HCI) by encouraging empirical research; promoting the use of knowledge and methods from the human sciences in design and evaluation of computer systems; promoting better understanding of the relation between formal design methods and system usability and acceptability; developing guidelines, models and methods by which designers may provide better human-oriented computer systems; and, cooperating with other groups, inside and outside IFIP, to promote user-orientation and humanization in system design. Thus, TC13 seeks to improve interactions between people and computers, encourage the growth of HCI research and disseminate these benefits world-wide.

The main orientation is toward users, especially the non-computer professional users, and how to improve human–computer relations. Areas of study include: the problems people have with computers; the impact on people in individual and organizational contexts; the determinants of utility, usability and acceptability; the appropriate allocation of tasks between computers and users; modelling the user to aid better system design; and harmonizing the computer to user characteristics and needs.

While the scope is thus set wide, with a tendency toward general principles rather than particular systems, it is recognized that progress will only be achieved through both general studies to advance theoretical understanding and specific studies on practical issues (e.g., interface design standards, software system consistency, documentation, appropriateness of alternative communication media, human factors guidelines for dialogue design, the problems of integrating multi-media systems to match system needs and organizational practices, etc.).

IFIP TC13 stimulates working events and activities through its working groups (WGs). WGs consist of HCI experts from many countries, who seek to expand knowledge and find solutions to HCI issues and concerns within their domains, as outlined below.

In 1999, TC13 initiated a special IFIP Award, the Brian Shackel Award, for the most outstanding contribution in the form of a refereed paper submitted to and delivered at each INTERACT. The award draws attention to the need for a comprehensive human-centered approach in the design and use of information technology in which the human and social implications have been taken into account. Since the process to decide the award takes place after papers are submitted for publication, the award is not identified in the proceedings.

WG13.1 (Education in HCI and HCI Curricula) aims to improve HCI education at all levels of higher education, coordinate and unite efforts to develop HCI curricula and promote HCI teaching.

WG13.2 (Methodology for User-Centered System Design) aims to foster research, dissemination of information and good practice in the methodical application of HCI to software engineering.

WG13.3 (HCI and Disability) aims to make HCI designers aware of the needs of people with disabilities and encourage development of information systems and tools permitting adaptation of interfaces to specific users.

WG13.4 (also WG2.7) (User Interface Engineering) investigates the nature, concepts and construction of user interfaces for software systems, using a framework for reasoning about interactive systems and an engineering model for developing user interfaces.

WG13.5 (Human Error, Safety and System Development) seeks a framework for studying human factors relating to systems failure, develops leading–edge techniques in hazard analysis and safety engineering of computer-based systems, and guides international accreditation activities for safety-critical systems.

WG13.6 (Human-Work Interaction Design) aims at establishing relationships between extensive empirical work-domain studies and HCI design. It promotes the use of knowledge, concepts, methods and techniques that enables user studies to procure a better apprehension of the complex interplay between individual, social and organizational contexts and thereby a better understanding of how and why people work in the ways that they do.

WG13.7 (Human–Computer Interaction and Visualization) is the newest of the working groups under the TC.13. It aims to establish a study and research program that combines both scientific work and practical applications in the fields of human–computer interaction and visualization. It integrates several additional aspects of further research areas, such as scientific visualization, data mining, information design, computer graphics, cognition sciences, perception theory, or psychology, into this approach.

New WGs are formed as areas of significance to HCI arise. Further information is available on the IFIP TC13 website: http://csmobile.upe.ac.za/ifip

IFIP TC13 Members

Australia
Judy Hammond
Australian Computer Society

Austria
Andreas Holzinger
Austrian Computer Society

Belgium
Monique Noirhomme-Fraiture
Federation des Associations Informatiques de Belgique

Brazil
Simone Diniz Junqueira Barbosa (TC 13 secretary)
Brazilian Computer Society (SBC)

Bulgaria
Kamelia Stefanova
Bulgarian Academy of Sciences

Canada
Heather O'Brian
Canadian Information Processing Society

China
Zhengjie Liu
Chinese Institute of Electronics

Cyprus
Panayiotis Zaphiris
Cyprus Computer Society

Czech Republic
Vaclav Matousek
Czech Society for Cybernetics and Informatics

Denmark
Annelise Mark Pejtersen
Danish Federation for Information Processing

Finland
Kari-Jouko Räihä
Finnish Information Processing Association

France
Philippe Palanque (TC 13 vice chair)
Societe des Electriciens et des Electroniciens (SEE)

Germany
Tom Gross
Gesellschaft fur Informatik

Hungary
Cecilia Sik Lanyi
John v. Neumann Computer Society (NJSZT)

Iceland
Marta Kristin Larusdottir
The Icelandic Society for Information Processing (ISIP)

India
Anirudha Joshi
Computer Society of India

Italy
Fabio Paternò
Italian Computer Society

Ireland
Liam J. Bannon
Irish Computer Society

Japan
Masaaki Kurosu
Information Processing Society of Japan

Kenya
Daniel Orwa Ochieng
Computer Society of Kenya

Malaysia
Chui Yin Wong
Malaysian National Computer Confederation

New Zealand
Mark Apperley
New Zeeland Computer Society (NZCS)

Nigeria
Chris Nwannenna
Nigeria Computer Society

Norway
Dag Svanes
Norwegian Computer Society

Poland
Juliusz L. Kulikowski
Poland Academy of Sciences

Portugal
Joaquim A. Jorge
Associação Portuguesa de Informática

Singapore
Henry Been-Lirn Duh
Singapore Computer Society

South Africa
Paula Kotzé
The Computer Society of South Africa

Spain
Julio Abascal
Asociación de Técnicos de Informática (ATI)

Sweden
Jan Gulliksen TC 13 (chair)
Swedish Interdisciplinary Society for Human–Computer Interaction (STIMDI) - Swedish Computer Society

Switzerland
Ute Klotz
Swiss Association for Research in Information Technology SARIT

The Netherlands
Gerrit C. van der Veer
Nederlands Genootschap voor Informatica

UK
Andrew Dearden
British Computer Society (BCS)

USA-based
John Karat
Association for Computing Machinery (ACM)
Nahum Gershon
The Computer Society, Institute of Electrical & Electronics Engineers (IEEE-CS)

Expert members
Nikos Avouris, *Greece*
Paula Kotzé, *South Africa*
Gitte Lindegaard, *Canada*
Annelise Mark Pejtersen, *Denmark*
Marco Winckler, *France*

Working Group Chairpersons

WG13.1 (Education in HCI and HCI Curricula)
Lars Oestreicher, *Sweden*

SIG13.1 (Interaction Design and International Development)
Janet Read, *UK*

WG13.2 (Methodology for User-Centered System Design)
Peter Forbrig, *Germany*

SIG13.2 (Interaction Design and Children)
Panos Markopoulous, *The Netherlands*

WG13.3 (HCI and Disability)
Gerard Weber, *Germany*

WG13.4 (joint with WG 2.7) (User Interface Engineering)
Fabio Paternó, *Italy*

WG13.5 (Human Error, Safety, and System Development)
Philippe Palanque, *France*

WG13.6 (Human-Work Interaction Design)
Torkil Clemmensen, *Denmark*

WG13.7 (Human–Computer Interaction and Visualization)
Achim Ebert, *Germany*

INTERACT 2011 Technical Committee

Conference Committee

General Co-chairs
Joaquim A. Jorge, *Portugal*
Philippe Palanque, *France*

Honorary Co-chairs
Larry Constantine, *Portugal*
Don Norman, *USA*
Anneliese Mark Pejtersen, *Denmark*

Technical Program Co-chairs
Daniel Gonçalves, *Portugal*
Nick Graham, *Canada*
Nuno Nunes, *Portugal*

Technical Program Committee

Demonstrations Co-chairs
Verónica Orvalho, *Portugal*
Greg Philips, *Canada*

Doctoral Consortium Co-chairs
Gitte Lindgaard, *Canada*
Manuel João Fonseca, *Portugal*

Full Papers Co-chairs
Nick Graham, *Canada*
Nuno Nunes, *Portugal*

Industrial Program Co-chairs
Antonio Câmara, *Portugal*
Miguel Dias, *Portugal*
Stacy Hobson, *USA*
Oscar Pastor, *Spain*
Virpi Roto, *Finland*

Interactive Posters Co-chairs
Adérito Marcos, *Portugal*
Monique Noirhomme-Fraiture, *Belgium*

Keynote Speakers Co-chairs
John Karat, *USA*
Jean Vanderdonckt, *Belgium*

Organization Overviews Co-chairs
Teresa Chambel, *Portugal*
Mary Czerwinski, *USA*

Panels Co-chairs
Regina Bernhaupt, *Austria*
Nuno Correia, *Portugal*
Peter Forbrig, *Germany*

Short Papers Co-chairs
Daniel Gonçalves, *Portugal*

Special Interest Groups (SIGs) Co-chairs
Gerrit van der Veer, *The Netherlands*
Teresa Romão, *Portugal*

Student Design Competition Co-chairs
Simone Diniz Junqueira Barbosa, *Brazil*
Luis Carriço, *Portugal*

Tutorials Co-chairs
José Creissac Campos, *Portugal*
Paula Kotzé, *South Africa*

Workshops Co-chairs
Julio Abascal, *Spain*
Nuno Guimarães, *Portugal*

Organizing Committee

Local Organization Co-chairs
Alfredo Ferreira, *Portugal*
Pauline Jepp, *Portugal*
Manuela Sado, *Portugal*

Multimedia Conferencing Co-chairs
José Coelho, *Portugal*
Lars Oestreicher, *Sweden*

Publications Co-chairs
Padro Campos, *Portugal*
Marco Winckler, *France*

Publicity Co-chairs
Paula Alexandra Silva, *Portugal*
Tiago Guerreiro, *Portugal*

Student Volunteers Co-chairs
Tiago Guerreiro, *Portugal*
Xavier Ferre, *Spain*
Effie Law, *UK*

Website Co-chairs
Alfredo Ferreira, *Portugal*

Associate Chairs - Full Papers

Julio Abascal, *Spain*
Jose Abdelnour-Nocera, *UK*
Silvia Abrahão, *Spain*
Vincent Aleven, *USA*
Nikolaos Avouris, *Greece*
Cecilia Baranauskas, *Brazil*
Simone Barbosa, *Brazil*
Patrick Baudisch, *Germany*
Regina Bernhaupt, *France*
Robert Biddle, *Canada*
Jeremy Birnholtz, *USA*
Kellogg Booth, *Canada*
Gaelle Calvary, *France*
Pedro Campos, *Portugal*
Torkil Clemmensen, *Denmark*
Nuno Correia, *Portugal*
Enrico Costanza, *UK*
Joelle Coutaz, *France*
José Creissac Campos, *Portugal*
Mary Czerwinski, *USA*
Peter Dannenmann, *Germany*
Andy Dearden, *UK*
Anke Dittmar, *Germany*
Ellen Do, *USA*
Gavin Doherty, *Ireland*
Andrew Duchowski, *USA*
Henry Been-Lim Duh, *Singapore*
Michael Feary, *USA*
Peter Forbrig, *Germany*
Nahum Gershon, *The Netherlands*
Marianne Graves Petersen, *Denmark*
Phil Gray, *UK*
Tom Gross, *Germany*
Mark D Gross, *USA*
Jan Gulliksen, *Sweden*
Michael Haller, *Austria*
Richard Harper, *UK*
Andreas Holzinger, *Austria*
Kasper Hornbaek, *Denmark*
Horst Hortner, *Austria*
Matt Jones, *UK*
Anirudha Joshi, *India*
Hermann Kaindl, *Austria*
Evangelos Karapanos, *Portugal*
Rick Kazman, *USA*
Ute Klotz, *Switzerland*
Vassilis Kostakos, *Portugal*
Masaaki Kurosu, *Austria*
Ed Lank, *Canada*
Marta Larusdottir, *Iceland*
Henry Lieberman, *USA*
Panos Markopolous, *The Netherlands*
Christian Muller, *Germany*
Miguel Nacenta, *Canada*
Laurence Nigay, *France*
Monique Noirhomme, *Belgium*
Eamonn O'Neill, *UK*
Ian Oakley, *Portugal*
Oscar Pastor, *Spain*
Fabio Paterno, *Italy*
Lia Patrício, *Portugal*
Helen Petrie, *UK*

Nitendra Rajput, *India*
Janet Read, *UK*
Dave Roberts, *UK*
Kari-Jouko Raiha, *Finland*
Miguel Sales Dias, *Portugal*
Jaime Sanchez, *Chile*
Robert St Amant, *USA*
Kamelia Stefanova, *Bulgaria*
James Stewart, *Canada*
Wolfgang Stuerzlinger, *UK*
Jan van den Bergh, *Belgium*
Gerrit van der Veer, *The Netherlands*
Jos van Leeuwen, *Portugal*
Gerhard Weber, *Germany*
Janet Wesson, *South Africa*
Marco Winckler, *France*
Volker Wulf, *Germany*

Associate Chairs - Short Papers

Jose Abdelnour-Nocera, *UK*
Elisabeth André, *Germany*
Mark Apperley, *New Zealand*
Nathalie Aquino, *Spain*
Simone Barbosa, *Brazil*
Alexander Boden, *Germany*
Gaelle Calvary, *France*
Robert Capra, *USA*
Luis Carriço, *Portugal*
Marc Cavazza, *UK*
Teresa Chambel, *Portugal*
Stéphane Conversy, *France*
Nuno Correia, *Portugal*
Tim Davis, *USA*
Antonella de Angeli, *UK*
Andy Dearden, *UK*
Anke Dittmar, *Germany*
Carlos Duarte, *Portugal*
Achim Eber, *Germany*
David Elsweiler, *UK*
Danyel Fisher, *USA*
Peter Forbrig, *Germany*
Tiago Guerreiro, *Portugal*
Jacek Gwizdka, *USA*
Marc Hassenzahl, *Germany*
Anirudha Joshi, *India*
Hermann Kaindl, *Austria*
Ute Klotz, *Switzerland*
Tessa Lau, *USA*
Gitte Lindgaard, *Canada*
Floyd Mueller, *USA*
Lennart Nacke, *Canada*
Yukiko Nakano, *Japan*
Monique Noirhomme, *Belgium*
Lars Oestreicher, *Sweden*
Eamonn O'Neill, *UK*
Dan Orwa, *Kenya*
Tim Paek, *USA*
Ignacio Panach, *Spain*
Fabio Paterno, *Italy*
Lia Patrício, *Portugal*
Nitendra Rajput, *India*
Francisco Rebelo, *Portugal*
Dave Roberts, *UK*
Teresa Romão, *Portugal*
Virpi Roto, *Finland*
Raquel Santos, *Portugal*
Beatriz Sousa Santos, *Portugal*
James Stewart, *Canada*
Sriram Subramanian, *UK*
Feng Tian, *China*
Manas Tungare, *USA*
Gerhard Weber, *Germany*
Astrid Weiss, *Austria*
Marco Winckler, *France*
Chui Yin Wong, *Malaysia*

Reviewers

Al Mahmud Abdullah, *The Netherlands*
Ana Paula Afonso, *Portugal*
Jason Alexander, *UK*
Jan Alexandersson, *Germany*
Dzmitry Aliakseyeu, *The Netherlands*
Majed Alshamari, *Saudi Arabia*
Margarita Anastassova, *France*
Craig Anslow, *New Zealand*
Caroline Appert, *France*
Nathalie Aquino, *Spain*
Pedro Arezes, *Portugal*
Ernesto Arroyo, *USA*
Mark Ashdown, *UK*
Ching man Au Yeung, *Japan*
Chris Baber, *UK*
Paula M. Bach, *USA*
Nilufar Baghaei, *New Zealand*
Sebastiano Bagnara, *Italy*
Gilles Bailly, *Germany*
Martina Balestra, *USA*
Emilia Barakova, *The Netherlands*
Jakob Bardram, *Denmark*
Shaowen Bardzell, *USA*
Javier Bargas-Avila, *Switzerland*
Louise Barkhuus, *Denmark*
Pippin Barr, *Denmark*
Barbara Rita Barricelli, *Italy*
Gil Barros, *Brazil*
Len Bass, *USA*
Remi Bastide, *France*
Rafael Bastos, *Portugal*
Eric Baumer, *USA*
Gordon Baxter, *UK*
Michel Beaudouin-Lafon, *France*
Nikolaus Bee, *Germany*
Yacine Bellik, *France*
Kawtar Benghazi, *Spain*
Mike Bennett, *USA*
François Bérard, *France*
Olav W. Bertelsen, *Denmark*
Nigel Bevan, *UK*
Ganesh Bhutkar, *India*
Matthew Bietz, *USA*
Mark Billinghurst, *New Zealand*
Dorrit Billman, *USA*
Fernando Birra, *Portugal*
Mike Blackstock, *Canada*
Marcus Bloice, *Austria*
Marco Blumendorf, *Germany*
Mark Blythe, *UK*
Cristian Bogdan, *Sweden*
Morten Bohoj, *Denmark*
Matthew Bolton, *USA*
Birgit Bomsdorf, *Germany*
Rodrigo Bonacin, *Brazil*
Sebastian Boring, *Canada*
Aviaja Borup, *Denmark*
Matt-Mouley Bouamrane, *UK*
Doug Bowman, *USA*
Giorgio Brajnik, *Italy*
Pedro Branco, *Portugal*
Willem-Paul Brinkman, *The Netherlands*
Gregor Broll, *Germany*
Christopher Brooks, *Canada*
Judith Brown, *Canada*
Steffen Budweg, *Germany*
Lucy Buykx, *UK*
Marina Buzzi, *Italy*
Daragh Byrne, *Ireland*
Cristina Cachero, *Spain*
Jeff Calcaterra, *USA*
Licia Calvi, *The Netherlands*
Eduardo Calvillo Gamez, *Mexico*
Maria-Dolores Cano, *Spain*
Xiang Cao, *China*
Cinzia Cappiello, *Italy*
Robert Capra, *USA*
Luis Carlos paschoarelli, *Brazil*
Stefan Carmien, *Spain*
Maria Beatriz Carmo, *Portugal*
António Carvalho Brito, *Portugal*
Luis Castro, *Mexico*
Daniel Cernea, *Germany*
Matthew Chalmers, *UK*

Teresa Chambel, *Portugal*
Beenish Chaudry, *USA*
Tao Chen, *China*
Fanny Chevalier, *Canada*
Keith Cheverst, *UK*
Yoram Chisik, *Portugal*
Yu-kwong Chiu, *China*
Georgios Christou, *Cyprus*
Andrea Civan Hartzler, *USA*
Laurence Claeys, *France*
Luis Coelho, *Portugal*
François Coldefy, *France*
Karin Coninx, *Belgium*
Maria Francesca Costabile, *Italy*
Céline Coutrix, *France*
Nadine Couture, *France*
Anna Cox, *UK*
David Coyle, *Ireland*
Leonardo Cunha de Miranda, *Portugal*
Edward Cutrell, *India*
Raimund Dachselt, *Germany*
José Danado, *Norway*
Tjerk de Greef, *The Netherlands*
Alexander De Luca, *Germany*
Luigi De Russis, *Italy*
Clarisse de Souza, *Brazil*
Alexandre Demeure, *France*
Charlie DeTar, *USA*
Ines Di Loreto, *Italy*
Eduardo Dias, *Portugal*
Paulo Dias, *Portugal*
Claire Diederich, *Belgium*
Andre Doucette, *Canada*
Carlos Duarte, *Portugal*
Emmanuel Dubois, *France*
Cathy Dudek, *Canada*
Andreas Duenser, *New Zealand*
Mark Dunlop, *UK*
Sophie Dupuy-Chessa, *France*
Matthew Easterday, *USA*
Achim Ebert, *Germany*
Florian Echtler, *USA*
Amnon Eden, *UK*
Serge Egelman, *USA*
Linda Elliott, *USA*
Niklas Elmqvist, *USA*
Alex Endert, *USA*
Dominik Ertl, *Austria*
Parisa Eslambolchilar, *UK*
Augusto Esteves, *Portugal*
Pedro Faria Lopes, *Portugal*
Robert Farrell, *USA*
Ian Fasel, *USA*
Ava Fatah gen. Schieck, *UK*
Jean-Daniel Fekete, *France*
Xavier Ferre, *Spain*
Mirko Fetter, *Germany*
Sebastian Feuerstack, *Brazil*
Nelson Figueiredo de Pinho, *Portugal*
George Fitzmaurice, *Canada*
Joan Fons, *Spain*
Manuel J. Fonseca, *Portugal*
Alain Forget, *Canada*
Florian Förster, *Austria*
Derek Foster, *UK*
Marcus Foth, *Australia*
Teresa Franqueira, *Portugal*
Mike Fraser, *UK*
Christopher Frauenberger, *UK*
André Freire, *UK*
Carla Freitas, *Brazil*
David Frohlich, *UK*
Dominic Furniss, *UK*
Luigi Gallo, *Italy*
Teresa Galvão, *Portugal*
Nestor Garay-Vitoria, *Spain*
Roberto García, *Spain*
Anant Bhaskar Garg, *India*
Vaibhav Garg, *USA*
Jose Luis Garrido, *Spain*
Nahum Gershon, *Canada*
Florian Geyer, *Germany*
Werner Geyer, *USA*
Giuseppe Ghiani, *Italy*
Andy Gimblett, *UK*
Patrick Girard, *France*
Sylvie Girard, *UK*
Leonardo Giusti, *Italy*
Guilherme Gomes, *Portugal*
Daniel Gonçalves, *Portugal*

José Luis González Sánchez, *Spain*
Phil Gosset, *UK*
Nitesh Goyal, *USA*
Toni Granollers, *Spain*
Anders Green, *Sweden*
Collin Green, *USA*
Saul Greenberg, *Canada*
Olivier Grisvard, *France*
Tiago Guerreiro, *Portugal*
Sean Gustafson, *Germany*
Mieke Haesen, *Belgium*
Jonna Häkkilä, *Finland*
Martin Halvey, *UK*
Judy Hammond, *Australia*
Mark Hancock, *Canada*
Morten Borup Harning, *Denmark*
John Harris, *Canada*
Kirstie Hawkey, *Canada*
Elaine Hayashi, *Brazil*
Brent Hecht, *USA*
Steffen Hedegaard, *Denmark*
Mathias Heilig, *Germany*
Ruediger Heimgaertner, *Germany*
Ingi Helgason, *UK*
Sarah Henderson, *New Zealand*
Bart Hengeveld, *The Netherlands*
Wilko Heuten, *Germany*
Michael Hildebrandt, *Norway*
Christina Hochleitner, *Austria*
Eve Hoggan, *Finland*
Paul Holleis, *Germany*
Clemens Holzmann, *Austria*
Jettie Hoonhout, *The Netherlands*
Michael Horn, *USA*
Eva Hornecker, *Germany*
Heiko Hornung, *Brazil*
Horst Hörtner, *Austria*
Juan Pablo Hourcade, *USA*
Aaron Houssian, *The Netherlands*
Andrew Howes, *UK*
Dalibor Hrg, *Germany*
Ko-Hsun Huang, *Portugal*
Jina Huh, *USA*
Tim Hussein, *Germany*
Dugald Hutchings, *USA*
Junko Ichino, *Japan*
Netta Iivari, *Finland*
Emilio Insfran, *Spain*
Samuel Inverso, *Australia*
Shamsi Iqbal, *USA*
Petra Isenberg, *France*
Howell Istance, *UK*
Linda Jackson, *USA*
Robert Jacob, *USA*
Mikkel Jakobsen, *Denmark*
Jacek Jankowski, *USA*
Hans-Christian Jetter, *Germany*
Sune Alstrup Johansen, *Denmark*
Jeff Johnson, *USA*
Simon Jones, *UK*
Martino Jose Mario, *Brazil*
Rui José, *Portugal*
Marko Jurmu, *Finland*
Don Kalar, *USA*
Vaiva Kalnikaite, *UK*
Martin Kaltenbrunner, *Austria*
Matthew Kam, *USA*
Mayur Karnik, *Portugal*
Hannu Karvonen, *Finland*
Sebastian Kassner, *Germany*
Dinesh Katre, *India*
Sevan Kavaldjian, *Austria*
Konstantinos Kazakos, *Australia*
Pramod Khambete, *India*
Vassilis-Javed Khan, *The Netherlands*
Hyungsin Kim, *USA*
Jayne Klenner-Moore, *USA*
Christian Kray, *UK*
Per Ola Kristensson, *UK*
Hannu Kukka, *Finland*
Andrew Kun, *USA*
H. Chad Lane, *USA*
Yann Laurillau, *France*
Effie Law, *Switzerland*
Marco Lazzari, *Italy*
Karin Leichtenstern, *Germany*
Juha Leino, *Finland*
Barbara Leporini, *Italy*
Sophie Lepreux, *France*
Olivier Lequenne, *France*

Chunyuan Liao, *USA*
Conor Linehan, *UK*
Agnes Lisowska Masson, *China*
Zhengjie Liu, *China*
Sara Ljungblad, *Sweden*
Claire Lobet, *Belgium*
Steffen Lohmann, *Spain*
Fernando Lopez-Colino, *Spain*
Anja Lorenz, *Germany*
Stephanie Ludi, *USA*
Bernd Ludwig, *Germany*
Andreas Luedtke, *Germany*
Jo Lumsden, *UK*
Kris Luyten, *Belgium*
Kent Lyons, *Canada*
Allan MacLean, *UK*
Joaquim Madeira, *Portugal*
Rui Madeira, *Portugal*
Angela Mahr, *Germany*
Stephann Makri, *UK*
Sylvain Malacria, *France*
Benjamin Mangold, *Germany*
Javier Marco, *Spain*
Gary Marsden, *South Africa*
Mark Marshall, *UK*
Hannah Marston, *Canada*
Jean-Bernard Martens, *The Netherlands*
Lynne Martin, *USA*
Diego Martínez, *Spain*
Célia Martinie, *France*
Masood Massodian, *New Zealand*
Sara Mastro, *USA*
Maristella Matera, *Italy*
Akhil Mathur, *Canada*
Eva Mayr, *Austria*
Davide Mazza, *Italy*
emanuela mazzone, *UK*
Gregor McEwan, *Canada*
Kevin McGee, *Singapore*
Marilyn McGee-Lennon, *UK*
Indrani Medhi, *India*
Gerrit Meixner, *Germany*
Guy Melancon, *France*
Eduarda Mendes Rodrigues, *Portugal*
Helena Mentis, *UK*
Tim Merritt, *Singapore*
Mei Miao, *Germany*
Alex Mitchell, *Singapore*
Robb Mitchell, *Denmark*
Jose Pascual Molina Masso, *Spain*
Francisco Montero, *Spain*
Meredith Morris, *USA*
Ann Morrison, *Denmark*
Christiane Moser, *Austria*
Omar Mubin, *The Netherlands*
Florian 'Floyd' Mueller, *USA*
Christian Mueller-Tomfelde, *Australia*
Michael Muller, *USA*
Maurice Mulvenna, *UK*
Dianne Murray, *UK*
Lennart Nacke, *Canada*
Peyman Nasirifard, *USA*
David Navarre, *France*
Ather Nawaz, *Denmark*
Luciana Nedel, *Brazil*
Vania Neris, *Brazil*
Colette Nicolle, *UK*
Femke Nijboer, *The Netherlands*
Valentina Nisi, *Portugal*
Leonel Nobrega, *Portugal*
Sylvie Noel, *Canada*
Manuel Noguera, *Spain*
Marianna Obrist, *Austria*
Johanna Renny Octavia, *Belgium*
Amy Ogan, *USA*
Michael O'Grady, *Ireland*
Kenton O'Hara, *UK*
Timo Ojala, *Finland*
Eugenio Oliveira, *Portugal*
Veronica Orvalho, *Portugal*
Nuno Otero, *Portugal*
Benoit Otjacques, *Luxembourg*
Ana Paiva, *Portugal*
Yue Pan, *USA*
Jose Ignacio Panach Navarrete, *Spain*
Alex Pang, *UK*
Nadia Pantidi, *UK*
Luca Paolino, *Italy*
Eleftherios Papachristos, *Greece*

Narcis Pares, *USA*
Andrew Patrick, *Canada*
Celeste Lyn Paul, *USA*
Sharoda Paul, *USA*
Andriy Pavlovych, *Canada*
Greg Phillips, *Canada*
Lara Piccolo, *Brazil*
Martin Pielot, *Germany*
Emmanuel Pietriga, *France*
franck poirier, *France*
Benjamin Poppinga, *Germany*
Christopher Power, *UK*
Raquel Prates, *Brazil*
John Precious, *UK*
Costin Pribeanu, *Romania*
Andreas Pusch, *France*
Alexandra Queirós, *Portugal*
Ismo Rakkolainen, *Finland*
Dave Randall, *UK*
Alberto Raposo, *Brazil*
Stuart Reeves, *UK*
Patrick Reignier, *France*
René Reiners, *Germany*
Malte Ressin, *UK*
Bernardo Reynolds, *Portugal*
Andy Ridge, *UK*
Xavier Righetti, *Switzerland*
Pierre Robillard, *Canada*
Simon Robinson, *UK*
Carsten Röcker, *Germany*
Yvonne Rogers, *UK*
Markus Rohde, *Germany*
Teresa Romão, *Portugal*
Virpi Roto, *Finland*
Anne Roudaut, *Germany*
jose rouillard, *France*
Mark Rouncefield, *UK*
Nicolas Roussel, *France*
Jaime Ruiz, *Canada*
Pascal Salembier, *France*
Antti Salovaara, *Finland*
Nithya Sambasivan, *USA*
Krystian Samp, *Ireland*
Paulo Sampaio, *Portugal*
Vagner Santana, *Italy*
Carmen Santoro, *Italy*
José Santos, *Portugal*
Teresa Sarmento, *Portugal*
Cheryl Savery, *Canada*
Dominique Scapin, *France*
Thomas Schlegel, *Germany*
Kevin Schneider, *Canada*
Johannes Schöning, *Germany*
Eric Schweikardt, *USA*
Gig Searle, *Austria*
Thomas Seifried, *Austria*
Marc Seissler, *Germany*
Malu Seixas, *Brazil*
Ted Selker, *USA*
Abi Sellen, *UK*
Dev Sen, *Canada*
Andrew Seniuk, *Canada*
Aaditeshwar Seth, *India*
Leslie Setlock, *USA*
Ehud Sharlin, *Canada*
Aditi Sharma, *South Africa*
Huihui Shi, *Germany*
Aubrey Shick, *USA*
Garth Shoemaker, *Canada*
Bruno Silva, *Brazil*
Frutuoso Silva, *Portugal*
Hugo Silva, *Portugal*
Klaus-Martin Simonic, *Austria*
Mikael B. Skov, *Denmark*
Roger Slack, *UK*
David Smith, *Canada*
Dustin Smith, *USA*
Thomas Smyth, *Canada*
William Soukoreff, *Canada*
Kenia Sousa, *Belgium*
Jan Stage, *Denmark*
Danae Stanton Fraser, *UK*
Gunnar Stevens, *Germany*
Erik Stolterman, *USA*
Markus Stolze, *Switzerland*
Steven Strachan, *USA*
Simone Stumpf, *UK*
Sriram Subramanian, *UK*
Ja-Young Sung, *USA*
Alistair Sutcliffe, *UK*

David Swallow, *UK*
Colin Swindells, *Canada*
Gerd Szwillus, *Germany*
Susanne Tak, *New Zealand*
Anthony Tang, *USA*
Charlotte Tang, *Canada*
Michael Tangermann, *Germany*
Franck Tarpin-Bernard, *France*
Alex Taylor, *UK*
Stephanie Teasley, *USA*
António Teixeira, *Portugal*
Michael Terry, *Canada*
VinhTuan Thai, *Ireland*
Harold Thimbleby, *UK*
Martin Tomitsch, *Australia*
Daniela Trevisan, *Brazil*
Sylvia Truman, *UK*
Manfred Tscheligi, *Austria*
Nikolaos Tselios, *Greece*
Simon Tucker, *UK*
Markku Turunen, *Finland*
Brygg Ullmer, *USA*
Leon Urbas, *Germany*
Teija Vainio, *Finland*
Leonel Valbom, *Portugal*
Egon L. van den Broek, *Austria*
Thea van der Geest, *The Netherlands*
Ielka van der Sluis, *Ireland*
Erik van der Spek, *The Netherlands*
Jean Vanderdonckt, *Belgium*
Radu-Daniel Vatavu, *Romania*
Manuel Veit, *France*
Jayant Venkatanathan, *Portugal*
Arnold P.O.S. Vermeeren, *The Netherlands*
Bart Vermeersch, *Belgium*
Jo Vermeulen, *Belgium*
Frédéric Vernier, *France*
Roel Vertegaal, *Canada*
Markel Vigo, *UK*
Nadine Vigouroux, *France*
Thomas Visser, *The Netherlands*
Stephen Voida, *USA*
Ivan Volosyak, *Germany*
Jade Wang, *USA*
Qing Wang, *China*
Leon Watts, *UK*
Astrid Weiss, *Austria*
Peter Wild, *UK*
Graham Wilson, *UK*
Max Wilson, *UK*
Heike Winschiers-Theophilus, *Namibia*
Jacob Wobbrock, *USA*
Peter Wolkerstorfer, *Austria*
Chui Yin Wong, *Malaysia*
Michael Wright, *UK*
Min Wu, *USA*
Peta Wyeth, *Australia*
Alvin W. Yeo, *Malaysia*
James Young, *Canada*
Ray Yun, *USA*
Loutfouz Zaman, *Canada*
Panayiotis Zaphiris, *Cyprus*
Martina Ziefle, *Germany*
Juergen Ziegler, *Germany*
Gottfried Zimmermann, *Germany*
Martin Zimmermann, *Germany*

Sponsors

Gold

Microsoft Research

Silver

hp

FCT
Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

Bronze

YDREAMS

Supporters

UNIVERSIDADE ABERTA
www.uab.pt

Organization

inesc id lisboa

apdsi
associação para a promoção e desenvolvimento da Sociedade da Informação

Table of Contents – Part IV

Long and Short Papers

Usable Privacy and Security

User Experience I

User Experience II

User Experience III

User Modelling

Visualization

Web Interaction

Demos

Doctoral Consortium

Industrial Papers

Interactive Posters

Organization Overviews

Panels

Special Interest Groups (SIGs)

Tutorials

Workshops

A Field Study of User Behavior and Perceptions in Smartcard Authentication

Celeste Lyn Paul[1], Emile Morse[2], Aiping Zhang[2], Yee-Yin Choong[2], and Mary Theofanos[2]

[1] University of Maryland Baltimore County, 1000 Hilltop Circle, Baltimore, Maryland, United States
cpaul2@umbc.edu
[2] National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, Maryland, United States
{emile.morse,aiping.zhang,yee-yin.choong, mary.theofanos}@nist.gov

Abstract. A field study of 24 participants over 10 weeks explored user behavior and perceptions in a smartcard authentication system. Ethnographic methods used to collect data included diaries, surveys, interviews, and field observations. We observed a number of issues users experienced while they integrated smartcards into their work processes, including forgetting smartcards in readers, forgetting to use smartcards to authenticate, and difficulty understanding digital signatures and encryption. The greatest perceived benefit was the use of an easy-to-remember PIN in replacement of complicated passwords. The greatest perceived drawback was the lack of smartcard-supported applications. Overall, most participants had a positive experience using smartcards for authentication. Perceptions were influenced by personal benefits experienced by participants rather than an increase in security.

Keywords: Human factors, multi-factor authentication, security, smartcard.

1 Introduction

In an attempt to protect their valuable data, organizations such as banking and financial institutions, corporations, and governments spend a large fraction of their information technology budgets on security [9]. Because of security, users of these systems must adhere to new business rules, information security policies, and training sessions – which takes time and effort away from their primary jobs.

There is a constant tension between users who are trying to get work done and security specialists who lock down systems to make them safer and more secure. Despite not being the security specialists, the burden of maintaining a secure work environment often falls on users. While the efficacy of such security practices can be debated [7], it is clear that there is a role for HCI researchers to play in understanding how users and security can be optimized to serve both. Users have often been described as the "weakest link" in information security, but they lack the proper motivation and tools to effectively contribute to the information security ecosystem

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 1–17, 2011.

[18]. Many approaches designed to increase security have placed additional burdens on users. Users now have longer, more complex, and more frequently changing passwords [23]; more restrictions on what they can do on the web [20], and they must use additional keys, tokens, or cards in order access information systems [2].

Homeland Security Presidential Directive 12 (HSPD12) [24] defines requirements for a standardized, U.S. Government-wide identification mechanism for a reliable identification credential. Personal Identity Verification (PIV) [15] is a smartcard-based multi-factor authentication (MFA) mechanism to increase security of government resources. The PIV smartcard contains personally identifiable information (PII) about the card holder such as his full name and agency, PKI certificates for authentication, encryption, and digital signature, and biometrics such as fingerprints and photo. It can be used for physical building access, for information system authentication, to support PKI, and as an identity card.

The PIV smartcard affects hundreds of thousands of U.S. Government employees and contractors. While the use of the smartcard is mandated by policy, we are concerned with how introducing this additional authentication factor will impact the perceptions, behaviors, and work processes of so many users. The PIV smartcards are meant to be used throughout the day as often as passwords would be used. This makes smartcard use very different from other types of card scenarios users may have experience with, e.g. weekly use of an Automatic Teller Machine (ATM) card. This is one of several reasons why we studied user behavior and perceptions using smartcards.

1.1 Research Goals

The purpose of this study was to understand factors that affect user behavior and perceptions in the use of smartcards for authentication and to examine factors that affect user behavior and perceptions of security in general. We had three main goals for the study. First, we wanted to learn how users would use the smartcards in their everyday work processes. Second, we wanted to learn how users' work processes might change to accommodate smartcard use. Third, we were interested in the user benefits and drawbacks of using smartcards in authentication.

2 Related Work

Multi-factor authentication (MFA) is the use of two or more independent security factors to authenticate to an information system. There are three factors commonly used in MFA [16]. The first factor is “something you know” such as a password, passphrase, or personal identification number (PIN). While passwords are perhaps the most common “something you know” authentication factor, challenges to password usability include the cognitive limits related to the number of passwords [6] and length of passwords [23] users must remember. A second factor is “something you are”, i.e. using biometrics such as fingerprints or a facial image. A benefit of using a biometric as an authentication factor is that it does not depend on secrecy [16]. It is something users will always have with them and will never need to remember. The third factor is “something you have” such as a key, token, or card. ATMs are a classic example of MFA using a card in combination with a PIN [26]. There are also additional authentication factors not as commonly used. For example, “someone you know” considers social network characteristics such as who you went to school with [4, 21]. Another factor, “something you do” considers behavioral characteristics such as online shopping habits [12].

In this paper, we focus on the “something you have” authentication factor, specifically the use of smartcards in MFA. Smartcards are tokens that include embedded chips that can store information. Smartcards are different from magnetic strip cards in that while both can store data, the chip in smartcards makes them more secure and provides more features. A smartcard can be used for MFA in several ways. The smartcard can be “something you have” by acting as a token. It can be used along with “something you know” such as a password or PIN. Smartcards can also support the “something you are” factor by storing biometric data on the embedded chip that can later be matched to the user.

Examples of smartcards used as authentication tokens similar to PIV include the European national electronic identity (e-ID) card [1] that stores different types of PII depending on the requirements of the issuing country. The U.S. Department of State uses the Biometric Logical Access Development and Execution (PKI/BLADE) card as an employee identity card and authentication token [8, 25]. The U.S. Department of Defense uses the Common Access Card as a military identity card, Geneva Conventions Identification Card, and authentication token [8].

While work has focused on smartcard security weaknesses such as problems with the embedded chip and PIN mechanism [14] and PIV implementation standard [11, 13], very little work has looked at smartcard usability [19]. Proctor et al. [17] compared multiple authentication methods through formal task analysis. They warn that the physical manipulation of a smartcard in the authentication process can add complexity to the authentication task and reduce ease of use compared to other authentication methods.

Braz and Robert [5] conducted a comparative analysis of different authentication methods. They compared methods such as passwords, smartcards, fingerprints, and keystroke patterns, on qualities such as benefits, drawbacks, security, and usability. Overall, they found that the smartcard rated as one of the most secure and usable methods for authentication.

Baldwin and Malone [3] described the use of smartcards in a health management system. The ability to store information, such as PII, increased the usefulness of smartcards for authentication beyond being a token. Patients identified themselves by presenting the smartcard and providing a PIN. Visits for care and therapy were recorded on the smartcard, creating a case history. Patients paid for services and filed claims using the smartcard. The smartcard provided an easy way to identify patients and help the patients manage their health care accounts.

The U.S. Department of State analyzed the impact of the PKI/BLADE smartcard with PIN on their userbase [25]. Their smartcard system allowed users to replace multiple username/password authentication credentials with a single smartcard/PIN credential. They analyzed their technical support logs to understand how the smartcard system affected user support requests. Before smartcard deployment, password reset support requests averaged 25.8% of all technical support requests per year. After smartcard deployment, password reset support requests dropped to an average of 10.6% of all technical support requests per year.

Strouble et al. [22] conducted a survey that looked at issues concerning security, usability, and productivity of smartcards. They found the use of a PIN instead of a password improved the security of the authentication mechanism, but did not necessarily increase usability of the smartcard system. Sixty-seven percent of the participants forgot their smartcard in a smartcard reader at least once, resulting in potential security risks. Six percent of those participants had their smartcard lost or stolen, resulting in security risks, replacement costs, and productivity loss.

3 Methodology

3.1 Participants

We studied 24 participants from a U.S. Government research institute over a period of approximately 10 weeks. There were 10 males, an average age of 47 (SEM ± 2), and and a distribution of education that was representative of the organization (8 high school degrees; 10 college degrees; 6 post-graduate degrees). Ten participants were engaged in technical work; five were support staff, e.g. secretarial work; and nine worked in an administrative specialty, e.g. finance. Two participants reported having experience with smartcards before the study.

Our study participants were recruited from an institution-wide technology pilot of 100 users testing the PIV smartcard technology. The technology pilot used the same smartcard system as our study; the only difference was that our study included additional research methodologies to assess the smartcard system. Recruitment in the technology pilot was designed to include test users from a representative sample of users in the institution, except for research scientists. Even without scientists, the technology pilot had a sampling of job roles and education similar to an industry corporation. Our study sample reflected the same demographics as the overall groups recruited in the technology pilot; except in our study there was a higher proportion of females. Participation in both the technology pilot and our study was voluntary. However, all participants were aware that the smartcard system would soon be mandatory for all users.

Fig. 1. From left to right: USB reader, laptop reader, integrated keyboard reader

3.2 Study Environment

Participants were given access to a fully functional PIV smartcard authentication system. Although we recruited from a "technology pilot," it was the system intended for institute-wide implementation. Participants had both an identity card and a smartcard. Participants also had both smartcards/PINs and usernames/passwords. The smartcard PINs were 6 to 8 numbers long and selected by participants at the time of smartcard issuance. The password length policy was 10 characters at the beginning of the study and changed to 12 characters one week after the study began. Passwords expired every 90 days and every participant changed passwords at least once during the study. Participants used their own computers and workspaces. Participants used either a laptop or desktop computer. Those with laptop computers docked the laptops at their workspaces and used external monitors, keyboards, and mice. Several participants with laptops occasionally worked from home, and several participants without laptops occasionally worked from home using their personal computers. All computers were running Windows XP with ActivClient smartcard middleware[1]. One of three types of smartcard readers was used by each participant (Fig. 1): an external USB reader (n=13), an internal laptop reader (n=6), or an integrated keyboard reader (n=5). Participant tasks were limited to those supported by the smartcard implementation. Supported tasks included using the smartcard to login/logout and to lock/unlock a computer, encrypt/decrypt and digitally sign an email or document, and authenticate to several smartcard-enabled web applications. In each case, participants could authenticate with either their smartcards/PINs or usernames/passwords.

3.3 Data Collection Methods

Several ethnographic research methods were used to obtain a breadth of coverage as well as a depth of understanding of our participants over the course of the study.

Periodic Surveys. Participants were asked to respond to the following statements in a standardized survey two or more times over the course of the study in order to evaluate their experiences with the smartcard system.

1. I am confident I know how the smartcard works and what it does.
2. I take the smartcard with me every time I leave my computer.
3. Using the PIN for the smartcard is easier than using a password.
4. The smartcard makes the login process easier than the current password-based login system.
5. The smartcard makes the login process faster than the current password-based login system.
6. Compared to using passwords, using the smartcard is more secure.
7. I [plan/will continue] to use the smartcard.
8. I would encourage my colleagues to switch to the smartcard.
9. I am [looking forward to/have] enjoyed using the smartcard

---

[1] Specific hardware and software products identified in this report were used in order to perform the evaluations described. In no case does such identification imply recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that the products and equipment identified are necessarily the best available for the purpose.

Table 1. Summary of study phases and data collection methods

Installation		**Smartcard Use**			**Wrap-up**
Pre-Install	*Post-Install*	*Daily*	*2-Week*	*6-Week*	*Exit*
Survey	Site visit Interview Survey	Daily diary Daily email	Site visit Interview Survey	Site visit Interview Survey	Site visit Interview Survey

Survey statements were rated on a 5-point scale from 'Strongly Disagree' (1) to 'Strongly Agree' (5) with 'Neither Agree nor Disagree' (3) as neutral. Participants were also given space for several questions to provide additional comments. While the statements were framed positively and may bias responses towards the positive, we analyzed the results in terms of relative change rather than absolute value.

Daily Diaries. Participants were asked to keep daily written diaries of notable smartcard events. We provided notebooks for them to write in and a guide on the types of events to note, such as forgetting to login using the smartcard. If participants were not comfortable writing in the provided notebook, they were encouraged to keep notes in an electronic document.

Daily Email Surveys. The daily email surveys asked participants to report about specific smartcard usage in a Yes/No format, such as "Did you use your password today?", and also provided an area for additional comments. The purpose of the daily email surveys was to supplement the daily diaries as a way of reporting critical events, and not as a quantitatively evaluated questionnaire.

Site Visits and Interviews. Site visits allowed us to observe smartcard use in the participant's natural environment. Interviews provided an opportunity to discuss participant's smartcard experience since the previous visit and review any critical events that were reported in the daily diaries or daily email surveys. Field notes from the site visit observations and interviews were recorded.

3.4 Procedure

The study was conducted in three phases: Smartcard Installation, Smartcard Use, and Study Wrap-Up. See Table 1 for a summary of research activities.

Smartcard Installation. Before we met with participants, we sent the Pre-Install survey and a technical support person provided a brief training document via email. Participants were asked to complete the Pre-Install survey before the first site visit. The training document contained a step-by-step scenario guide on how to complete smartcard-related activities. The technical support person then installed the smartcard hardware and software, demonstrated how to use the smartcard, and guided participants through several smartcard use scenarios. The support person also asked participants to perform tasks, such as locking and unlocking their computers. After installation was complete, we conducted the Post-Install site visit. We observed participants using their smartcards, interviewed them about their first time experiences with the smartcards, administered the Post-Install surveys, and provided them the Daily Diaries and guide.

Smartcard Use. During smartcard use, participants were asked to keep diaries of notable smartcard usage and events. The daily email survey was sent near the end of the day for participants to complete and return. Researchers would review the daily diaries and email surveys before each site visit in order to discuss any critical events during the interview, if necessary. Site visits were conducted two and six weeks after Smartcard Installation. During these site visits the researchers observed participants using their smartcards, interviewed them about their smartcard experiences to date, and administered the 2-week or 6-week surveys.

Study Wrap-up. Study wrap-up activities took place between 10 to 12 weeks after Smartcard Installation. Extended business travel or paid time off was not counted towards participants' total study time. This resulted in slightly longer periods of participation for a few participants. During the last site visit, we observed participants using their smartcards, interviewed them about their overall smartcard experiences, administered the Exit surveys, and collected the Daily Diaries.

Table 2. Survey statement with mean values and 95% confidence intervals

Periodic Survey Statement				
Pre-Install	*Post-Install*	*2-Week*	*6-Week*	*Exit*
1. User confidence with smartcard				
3.50 ± 0.38	4.42 ± 0.19	4.33 ± 0.30	4.12 ± 0.40	4.20 ± 0.34
2. Remembering smartcard				
–	–	3.67 ± 0.47	4.04 ± 0.43	3.88 ± 0.41
3. PIN is easier than password				
3.5 ± 0.29	3.92 ± 0.38	4.17 ± 0.28	4.24 ± 0.38	4.32 ± 0.39
4. Smartcard easier than password				
3.40 ± 0.34	3.73 ± 0.39	4.08 ± 0.35	4.24 ± 0.31	4.24 ± 0.31
5. Smartcard faster than password				
3.33 ± 0.33	3.50 ± 0.41	3.83 ± 0.40	3.68 ± 0.52	3.64 ± 0.49
6. Smartcard more secure than password				
3.73 ± 0.34	–	–	3.60 ± 0.32	3.84 ± 0.29
7. Smartcard adoption				
4.19 ± 0.19	4.35 ± 0.19	4.29 ± 0.38	–	4.48 ± 0.32
8. Recommend smartcard to colleagues				
–	–	4.08 ± 0.33	4.04 ± 0.26	4.32 ± 0.31
9. Smartcard satisfaction				
3.92 ± 0.26	–	–	–	4.32 ± 0.29

4 Results

We report the analysis of our study results in three sections. First, we provide descriptive statistics for the quantitative data collection methods. Second, we discuss our analysis of the qualitative data, including in-context discussion of the quantitative results. Third, we reflect on the methodology and provide lessons learned.

4.1 Descriptive Statistics of Quantitative Methods

Periodic Surveys. Results from the periodic surveys are reported as mean values with 95% confidence intervals in Table 2. Pairwise comparisons are discussed alongside the qualitative results. The Wilcoxon signed-rank test is used for pairwise comparisons and Kendall's correlation is used for measuring relationships.

Daily Email Surveys. There were a total of 682 daily email surveys collected; the average number of emails collected from each participant during the study for each participant was 28.4 (SEM ±2.5). The number of email surveys collected diminished over the course of the study with 3.5 (SEM ± 0.2) surveys/week collected between the Post-Install and 2-Week site visits, 2.5 (SEM ± 0.3) surveys/week between the 2-Week and 6-Week site visits, and 2.3 (SEM ± 0.3) surveys/week between the 6-Week and Exit site visits.

4.2 Analysis of Observations

Table 3 provides a summary of reported issues with discussion in the following section.

User Confidence. Participant confidence in using the new smartcard authentication system increased immediately after installation (Pre-install/Post-install: $W=3.87$, $p=0.001$) and remained high throughout the study (Post-install/Exit: Kendall's $W=0.04$, $p=0.41$).

Seventeen participants reported reading the provided training materials before the smartcard hardware and software were installed. Several participants indicated that they preferred in-person training rather than reading the training documentation. One participant noted that he preferred a "*hand holding demo*" (P3) when using a new system for the first time. Another participant indicated that she "*learn[s] better by hands on training*" (P12). The personalized training could be a factor in the significant increase of confidence post-installation. However, not all participants felt the need for attended training. As one participant explained, "*I prefer to jump in and just start using any new product, referring to the documentation only when I get stuck or find features I'm curious about*" (P16).

Smartcard Readers. Fifteen participants used USB readers that were placed on their desks. The location of a reader on a desk varied as did the number of items that might obscure it from view. The reader itself added to clutter on the desk, as one participant commented, "*I know I'm going to dislike the wire connecting the smartcard reader to the computer – makes for a messy desk!*" (P7). This participant attempted to clean her desk by moving the reader out of the way, but later attributed the position of the reader to why she may have forgotten her smartcard, "*I tidied [the] smartcard reader cord - made [the] reader less intrusive, but moved it further out of [the] workspace. It may be a reason for forgetting to remove the smartcard to lock my computer.*"

Another participant also blamed his USB reader for why he forgot his card, "*I walked away at one point and forgot my smartcard. This has happened once or twice, and it makes me think that smartcard readers should probably be fairly visible*" (P14). Later that week, he tried a keyboard reader with success, "*Switched out my [USB] reader & keyboard for a new keyboard that included a smartcard reader. I like the setup much better. Less clunky, and the smartcard is more visible.*" Besides adding to clutter on the desk, an additional problem with the USB reader was that it was not attached to a stable object and required participants to use two hands when removing and inserting the smartcard. One participant remedied this problem by attaching her USB reader to her computer with rubber bands.

Table 3. Summary of study observations by issue topic

Issue Topic	Observation
User Confidence	• Confidence in using smartcards increased after Installation
Smartcard Reader	• New object in environment to get used to • Reason for forgetting smartcard in reader • Form factor may matter
Using Smartcards	• Smartcards easier for authentication than passwords • Forgot to remove smartcards from readers • Forgot to use smartcards to login • Forgot to use smartcards to lock screens • Forgot to use smartcards to unlock screens • Smartcard slower to login than password but faster otherwise • Unattended smartcard message is sometimes useful
Password vs. PIN	• PINs easier to use than passwords • Password requirements were burdensome • Passwords became difficult to remember because of smartcard use • Various password management strategies
Certificates	• Selecting certificate for web application authentication was confusing • Certificates could not be backed up or transferred
Digital Signatures and Encryption	• Digital signatures and encryption were easy to use • Did not understand digital signatures and none would use them • Understood encryption but few would use it • Implementation does not support inter-institutional use
Security Behavior	• Smartcards gave perceived increase in security • Low understanding of how or why security works • PII users were the most security conscious
Overall Experience	• Overall positive experience with smartcards • Most would recommend smartcards to colleagues • Most would continue using it voluntarily • Some had problems fitting smartcards into work processes

Not all participants could use a keyboard reader because their keyboards were in keyboard trays attached to their desks. Several of these participants who used keyboard trays also kept their keyboards partially hidden under their desks while they typed. Unless the keyboard was completely pulled from under the desk, it would not fit under the desk with the smartcard in the keyboard reader.

Using Smartcards. Remembering to remove their smartcards from their readers was a commonly reported incident by participants. Thirteen participants (54%) forgot their smartcards in their readers at least once during the study. The most common scenario for participants to forget their smartcards was during short trips out of their work areas, such as down the hall to visit a colleague or visit the restroom. Three participants forgot their smartcards in their readers after leaving an access-controlled area, and had to rely on their non-smartcard identity cards to gain access to their buildings. Six of 24 participants forgot their smartcards in their readers overnight. One participant forgot her smartcard in the reader overnight and drove back to campus to retrieve the card. Three participants reported forgetting their smartcards at home and had to use their passwords to login. Incidents where participants left their smartcards in the readers overnight or at home only occurred once or twice per participant. Even though half of the participants reported in interviews that they forgot their smartcards in the readers at some point during the study, most participants reported remembering their smartcards most of the time by the end of the study. Many participants who forgot their smartcards in readers early in the study reported that they forgot their smartcards less often as time went on. A few of these participants pointed out that it seemed to take them about one month before they developed a habit for using and remembering their smartcards; however, this change is not indicated in the periodic survey results. One participant described a system she developed to help her remember her smartcard; when she removed her smartcard from her badge holder, she would place the badge holder in front of her keyboard. It served as a reminder for her to take her smartcard before she left the office. This participant did not report forgetting her smartcard at any time during the study.

There were several reasons why participants did not use the smartcard to login or lock their computers. In the beginning of the study, most participants simply forgot to use their smartcards because it was not yet a habit. This was especially true for participants who had good security habits, such as those who consistently locked their computers with the keyboard when they left their workspace. As one participant stated, "This is going to take some getting used to - I have been using the keyboard to lock my machine for 10 years - hard habit to break" (P21). Other reasons participants did not use the smartcard to login or unlock their computers included because they forgot their smartcard at home, were using multiple computers at once, or were prompted to enter a username and password by the software. The design of the login dialog may have contributed to whether participants used their usernames and passwords to login. By default, the computer displayed a username and password dialog instead of a PIN dialog. A few participants discovered if they pressed Escape on the keyboard with their smartcards in their readers, a PIN dialog would display. This information was shared with the other participants. If participants did not use the smartcard to login to a session, they would not be able to remove the smartcard from the reader in order to quickly lock the computer. This caused some confusion in the beginning of the study when participants were not yet consistently using the smartcard to login, “*After logging in with my keyboard, I locked the machine but the smartcard could not unlock it until I logged in and locked the machine again*” (P14).

Unlocking computers also caused confusion for several participants in the beginning of the study. When returning to their locked computers, out of habit these participants would use CTRL+ALT+DEL in order to cancel the screensaver and unlock their computers. Using this key combination displayed the username and password dialog. Since participants were prompted with username and password dialogs, they entered their usernames and passwords and created sessions that could not be locked by removing their smartcards. It took time and practice for these participants to get used to using their smartcards to unlock their computers without pressing CTRL+ALT+DEL.

Participants were neutral whether smartcard authentication was faster than passwords (Exit: 3.64 ± 0.49). Nine participants noticed that using smartcards took longer to authenticate and login than using usernames and passwords. Smartcard login was observed by participants to take 10-30 seconds longer than their password logins. The physical act of inserting the card also added time to the login process. While the smartcard is slower in some cases, most participants considered the overall system tradeoffs and still felt smartcards were faster and easier to use, “*Unlocking when I returned to my desk was simple and no harder or time consuming than username and password – maybe easier*” (P7).

Participants who worked both at their computers and elsewhere in their work areas often experienced automatic computer screen locking after 15 minutes of inactivity. When the screen automatically locked with a smartcard in the reader, a message describing an unattended smartcard appeared. Participants who frequently worked at their work areas found these error messages frustrating, “*It's not unattended, I'm right here!*” (P17). However, participants who accidentally locked their computers using their keyboards felt the unattended message was useful, “*The message helped me not forget my smartcard when I accidentally locked using the keyboard*” (P2).

Passwords vs. PINs. Overall, participants found using the smartcard easier than using passwords (Exit, 4.24 ± 0.31). Participants also found that logging in with the PIN was easier than using their passwords (Exit, 4.32 ± 0.39). Many participants noted that the PIN was an important feature of their positive smartcard experience, particularly for its ease-of-use. The PIN was numeric-only while the system password consisted of numbers, upper- and lower-case letters, and special characters. The PIN never expired compared to the system password that expired every 90 days. The PIN was six to eight characters in length compared to the system password requirements of 10 or 12 characters in length. One participant noted the importance of the PIN not changing, “*If the PIN has to be changed as often as the password, there would be a reduced benefit to having the PIN*” (P21). Password length had a noticeable effect on participants' perceptions. Several participants complained how the new 12-character password requirement made it more difficult to remember their passwords.

The password length requirement was not the only burden passwords placed on participants. Smartcard use prevented participants from practicing their passwords as often as before the study. Several participants felt they were at risk for forgetting their passwords, “*It is an effort remembering my system password*” (P15). Some participants needed their primary network password to login to computers that were not smartcard-enabled, providing an opportunity to practice their passwords if they were synced. However, participants who did not sync passwords or used their passwords to only login to their computers were at a greater risk for forgetting.

Participants also described various ways they managed passwords before their experience with smartcards. Nine participants reported managing their passwords by recording them on paper and storing them in their wallets, purses, or drawer in their offices. Some participants also attempted syncing passwords for multiple applications so they had fewer passwords to remember. However, not all password requirements were the same and it was easy for their passwords to get out of sync. It was also a hassle to retrieve a password for every account. Two participants reported using software to save and manage passwords. Some participants anticipated the smartcard moving towards a SSO solution, "*The idea of having one "pin" for all applications is a dream come true! Also - less work for both user and the IT help desks for resetting passwords!*" (P18).

Certificates. Authentication was supported for several web applications. Participants authenticated to web applications by visiting the login page. A browser dialog appeared asking participants to select a certificate to use for authentication. For web application authentication, the certificate used for authentication did not matter. However, the authentication process was different depending on which certificate participants choose. If participants chose the non-repudiation certificate used for digital signatures, they were asked to enter a PIN. If participants chose the certificate used for encryption, they were not asked to enter a PIN. However, without expert knowledge in certificates it was difficult for participants to identify the encryption certificate from the digital signature certificate.

The smartcard authentication system also does not allow certificates to be backed up or saved/transferred. Three participants expressed concerns about the lifetime of certificates used to encrypt email and documents. If a smartcard is lost, stolen, expires, or is replaced, the certificates are lost forever. Previously encrypted email and documents could no longer be decrypted and the information would no longer be accessible. One of the participants worked with financial information and was concerned with not being able to access old encrypted data because his data needs to be available for auditing.

Digital Signatures and Encryption. The most infrequently used smartcard features were digital signatures and encryption. Once familiar with the functionality, participants were comfortable with digitally signing and encrypting email and documents and found both easy to use. However, most participants did not know when they would have a need to sign or encrypt an email or document, even after reading the sample use cases in the training document and after discussions with researchers during interviews. The training documentation explained how to sign and encrypt, but not why a participant would want to do so. No participants indicated a need for signing email or documents; although, several participants routinely tested the features. Several participants stated they would not consider using signing and encrypting unless it became policy, "*I see no value in a digitally signed email and would do so only if I was required to*" (P23). Two participants tried encryption to send passwords through email, and found it useful. The few participants who considered using signing and encrypting were those who already used some type of signing and encrypting in other applications.

All participants unintentionally signed email at some point in the study due to a technical problem that temporarily changed an option in their email clients. While most participants immediately noticed a difference in behavior and found typing a PIN for every sent email inconvenient, one participant decided to experiment with email signing for the rest of the study. This participant did not find digital signatures a huge burden, but he was also a technical user who understood the purpose of digital signatures. He also acknowledged the functionality might not be for everyone, "*I send a small number of emails on a typical day, so it isn't a big deal for me, but if I had to enter [the PIN] 50 or 100 times a day, it would become bothersome*" (P21). Few participants could describe a practical use for digital signatures. As one participant expressed his doubts about the usefulness of digital signatures, "*I don't see the point. People are going to know who I am based on what I say in the email*" (P2).

Participants who worked regularly with PII considered the possibility of signing and encrypting email or documents, but in practice found it impossible to use with the current smartcard implementation. PII in the workplace was shared between coworkers through secure applications, shared files on a shared remote storage location, or paper. Sharing PII out of the workplace with external contacts was the most common scenario where encryption would be useful. However, a participant could not encrypt a message to another user or verify that user's signature unless they had the user's public certificate. There was no infrastructure to easily obtain, share, and verify certificates from contacts outside the institute. Several participants explained how they thought encryption would be more useful once they knew their colleagues outside the institution were setup and supported for sharing.

Security Behavior. Many participants commented that they felt the smartcard was more secure; however, reasons why they felt the smartcard was more secure varied. For some participants, using the smartcard "*enforces good habits*" (P14) and encouraged participants to lock their computers. As one participant described, "*I felt my computer was more secure than ever before because I was forced to secure my computer each time I left my office by taking my smartcard with me each time*" (P15). At the same time, one participant who had already developed good computer locking habits was afraid that the smartcard had negatively impacted how often she locked her computer in the beginning of the study.

While participants felt the smartcards were more secure, few could articulate how or why. Three participants explained the smartcard was an additional security factor. Two participants noted the smartcards increased security because they would be difficult to crack or copy. There were mixed feelings about the use of a PIN instead of a password. Some participants felt that the shorter PIN was a benefit because the PIN is easy to remember and security would increase because it would not need to be written down. However, one participant was concerned that the PIN was not complex or long enough and might pose a security risk, "*The PIN for the smartcard is all numeric & 6-8 digits. Not sure if the multi-factor aspect makes it more secure than 1 more complex password alone*" (P23).

At the beginning of the study, several participants described themselves as being very diligent about security. Participants who seemed to have the best security habits, such as consistently locking their computer screens when leaving the office, were those who worked with PII or financial data. These participants were very aware of the sensitivity of the information they worked with, and felt that most security measures were justified. Participants who did not share these job roles had very different attitudes toward security. There seemed to be a high amount of inter-office trust, i.e. coworkers were not the threat. Two participants indicated they left their smartcards near their readers when they temporarily left their workspaces. One of these participants attempted to justify this behavior, "*It is OK since no one can get to my computer without the PIN and my other card can get me in the building.*" (P18). Although this participant was warned that her non-smartcard identity card would be phased out, she did not consider this when she developed this behavior.

Overall Experience. Even though each participant reported at least one problem or issue regarding smartcard use during the study the overall satisfaction of participants at the end of the study seemed positive. All but 3 participants (88%) indicated during the exit interview that they would recommend the smartcard to their colleagues. In general, participants were positive about using the smartcard, especially those in administrative job roles. These participants used the smartcards to access multiple applications and described a noticeable benefit.

Overall, participants were very positive about continuing their smartcard use after the study (Exit, 4.48 ± 0.32). However, not all participants had a consistently positive experience with the smartcards. Bad experiences and general frustration with the smartcards seemed to have an effect on some participants' behavior and perceptions. While minor problems, especially at the beginning of the study, were expected and accepted by most participants, issues that were persistent and affected work productivity were not acceptable. Early frustration with the smartcard had noticeable effects, "*Off to a bad start today and never fully recovered. I didn't use the smartcard for most of the day*" (P7). Although this participant had a particularly frustrating day, she resumed using the smartcard the next day and reported positive comments about her smartcard experience during the exit interview.

Another participant could not find a way to fit smartcard use into her existing work process. The smartcard authentication was noticeably slow to her and she described being "*always in a hurry to login*" (P24). She explained in the exit interview that if the smartcard became policy, she would use it; however, there were not enough benefits to encourage her to continue using the smartcard voluntarily. This benefit tradeoff was discussed by another participant who shared the same sentiments about recommending the technology, "*I can't really recommend it, as it has few clear benefits to offset the downsides*" (P23).

4.3 Methodology Lessons

This study utilized a number of qualitative field methods including daily diaries, daily surveys, periodic surveys, periodic user interviews, and field observations. The lessons learned in this study will help improve the design and methods of future work.

Most participants kept the written diary until the 2-week interview, but few participants wrote much after that. Several participants kept a written journal in a text file on their computers because they felt it was easier to keep than a written diary. Most participants summarized daily diary events in the daily surveys. Despite the lack of participation in keeping the daily diaries, we feel that providing the diaries helped participants understand the purpose and type of information we were interested in. Although the study designs were different, our methodology experiences are consistent with those reported in a diary study of password use conducted by Inglesant and Sasse [10].

As previously discussed, daily surveys were answered more frequently by participants in the beginning of the study than at the end of the study. When asked during interviews, several participants reported not responding to daily surveys because they did not have new events or comments to report and felt their responses would not add value to their participation. Modifying daily survey questions to fit with participants' evolving user experience could help encourage continued daily survey participation.

We observed participants' smartcard use from installation through 10 weeks of use. We observed new behaviors and perceptions up to six weeks after installation. After six weeks up until 10 weeks of observation, the rate of new observations was infrequent. At a minimum, we recommend users be studied for six weeks to be able to observe any transitions from learning to everyday use. Observations for less than two weeks after installation would not be sufficient to fully understand the impact the new technology has on participant behavior and perceptions.

5 Summary

Over the course of 10 weeks, we studied how 24 participants used smartcards in multi-factor authentication. We had three main goals for the study. First, we wanted to learn how users would use the smartcards in their everyday work processes. All participants quickly became confident using the smartcards and reported them easier to use than password authentication. One consistent problem was that many participants forgot to use their smartcards to login, lock, or unlock their computers during the beginning of the study. One reason was because they had not yet developed a habit for using their smartcards; another reason was that the form factor of the smartcard reader also had an effect on participants forgetting their smartcards.

Second, we wanted to learn how users' work processes might change to accommodate smartcard use. We found two ways smartcards changed the way some of our participants worked. First, the physical presence of new hardware to interact with had an effect on user behavior with the smartcard system. Some participants felt the location of their smartcard readers was the cause of forgetting their smartcards in the readers. They experimented with the location and form factor of their smartcard readers until they were satisfied with their setup. We also describe one participant who developed a novel strategy to help her consistently remember her smartcard. Second, digital signatures and encryption added an extra element to participants' document and email usage. Several participants experimented with digital signatures and encryption to see how they could incorporate the functionality into their work processes. However, few participants had a need for digital signatures and encryption.

Third, we were interested in the user benefits and drawbacks of using smartcards in authentication. The greatest perceived benefit of using smartcards was the use of a PIN in replacement of a password. Participants immediately noticed the difference between a shorter numeric PIN compared to the longer alpha-numeric-special character-password they were used to. Participants also appreciated that the same PIN was used for authenticating to multiple accounts, alleviating the need for complex password management practices. The greatest drawback to using smartcards was the lack of authentication support for more applications. The few participants who had trouble integrating the smartcard into their work processes used the smartcard only to authenticate to their computers. These participants experienced more overhead for smartcard use with fewer benefits than the participants who used the smartcards to authenticate to multiple systems.

Overall, participants had a positive experience with the smartcards and most indicated they would continue using the smartcards voluntarily and would even recommend using smartcards to their colleagues. It was interesting to see that participants' perceptions of the value of smartcards were related to personal benefits gained, such as an alternative to managing passwords, over the institutional benefit from increased security of a multi-factor authentication method.

Our multi-method approach provided us a richer understanding of smartcard use that could not be attained through traditional laboratory testing. While each method contributed to the understanding of our participants' behavior and perceptions, our experience provided lessons to further improve the methods for future or similar work.

Acknowledgements. We would like to thank Serge Egelman for his comments.

References

1. Arora, S.: National e-ID card schemes: A European overview. Information Security Technical Report 13(2), 46–53 (2008)
2. Aussel, J.: Smartcards and Digital Security. Computer Network Security 1, 42–56 (2007)
3. Baldwin, M.K., Malone, B.M.: Utilizing Smart Cards for Authentication and Compliance Tracking in a Diabetes Case Management System. In: Proceedings of ACM Conference on Software Engineering, pp. 521–522 (2008)
4. Brainard, J., Juels, A., Rivest, R.L., Szydlo, M., Yung, M.: Fourth-Factor Authentication: Somebody You Know. In: Proceedings of ACM CCS, pp. 168–178 (2006)
5. Braz, C., Robert, J.M.: Security and Usability: The Case of the User Authentication Methods. In: Proceedings of d'Interaction Homme-Machine, pp. 199–203 (2006)
6. Florencio, D., Herley, C.: A large-scale study of web password habits. In: Proceedings of ACM Conference on the World Wide Web, pp. 657–666 (2007)
7. Herley, C.: So Long and No Thanks for the Externalities: The Rational Rejection of Security Advice by Users. In: Proceedings of New Security Perspectives Workshop 2009 (2009)
8. Identity, Credential and Access Management Subcommittee.: The Realized Value of the Federal Public Key Infrastructure (FPKI) v1.0.0. January 29 (2010), http://www.idmanagement.gov/

9. Information Technology Sector Coordinating Council.: Response to White House Cyber Review Questions. ITSCC March 20 (2009), http://www.it-scc.org/documents/itscc/ITSCCandCommunicationsSCCJointResponsetotheWhiteHouseCyberspacePolicyReview_3_20_2009.pdf
10. Inglesant, P.G., Sasse, M.A.: The true cost of unusable password policies: password use in the wild. In: Proceedings of ACM Conference on Computer-Human Interaction, pp. 383–392 (2010)
11. Irwin, C.S., Taylor, D.C.: Identity, Credential, and Access Management at NASA, from Zachman to Attributes. In: Proceedings of IDtrust 2009, pp. 1–14 (2009)
12. Jakobsson, M., Shi, E., Golle, P., Chow, R.: Implicit Authentication for Mobile Devices. In: Proceedings of USENIX Workshop on HotSec (2009)
13. Karger, P.A.: Privacy and Security Threat Analysis of the Federal Employee Personal Identity Verification (PIV) Program. In: Proceedings of the Symposium on Usable Privacy and Security 2006, pp. 114–121 (2006)
14. Murdoch, S.J., Drimer, S., Anderson, R., Bond, M.: Chip and PIN is Broken. In: Proceedings of IEEE Symposium on Security & Privacy 2010, pp. 433–446 (2010)
15. National Institute of Standards and Technology: Personal identity verification (PIV) for federal employees and contractors. FIPS PUB 201-1 (2006)
16. O'Gorman, L.: Comparing Passwords, Tokens, and Biometrics for User Authentication. Proc. IEEE 91(12), 2019–2040 (2003)
17. Proctor, R.W., Lien, M.C., Salvendy, G., Schultz, E.E.: A Task Analysis of Usability in Third-Party Authentication. Information Security Bulletin 5(3), 49–56 (2000)
18. Sasse, M.A., Brostoff, S., Weirich, D.: Transforming the 'weakest link' – a human/computer interaction approach to usable and effective security. BT Technology Journal 19(3), 122–131 (2001)
19. Sasse, M.A.: Usability and Trust in Information Systems. In: Cyber Trust & Crime Prevention Project. University College, London (2004)
20. Schechter, S.E., Dhamija, R., Ozment, A., Fischer, I.: The Emperor's New Security Indicators. In: Proceedings of IEEE Symposium on Security & Privacy 2007, pp. 51–65 (2007)
21. Schechter, S., Egelman, S., Reeder, R.W.: It's not what you know, but who you know: a social approach to last-resort authentication. In: Proc. ACM CHI 2009, pp. 1983–1992 (2009)
22. Strouble, D.D., Schechtman, G.M., Alsop, A.S.: Productivity and Usability Effects of Using a Two-Factor Security System. In: Proceedings of SAIS, pp. 196–201 (2009)
23. Summers, W.C., Bosworth, E.: Password policy: the good, the bad, and the ugly. In: Proceedings of WISICT 2004, pp. 1–6 (2004)
24. U.S. Department of Homeland Security: Policy for a common identification standard for federal employees and contractors. Homeland Security Presidential Directive HSPD-12, August 27 (2004)
25. U.S. Department of State: Cost/Benefit Comparison between PKI/BLADE and Password-based Authentication v1.0, July 2010 (2010), Point of contact, Steven Gregory, gregoryse@state.gov
26. Weir, C.S., Douglas, G., Richardson, T., Jack, M.: Usable security: User preferences for authentication methods in eBanking and the effects of experience. Interacting with Computers 22(3), 153–164 (2010)

Improving Computer Security Dialogs

Cristian Bravo-Lillo[1], Lorrie Faith Cranor[1,3], Julie Downs[2], Saranga Komanduri[3], and Manya Sleeper[3]

[1] Engineering and Public Policy,
[2] Social and Decision Sciences,
[3] Computer Science,
Carnegie Mellon University, Pennsylvania, USA
{cbravo,lorrie,downs,msleeper,sarangak}@cmu.edu

Abstract. Security dialogs warn users about security threats on their computers; however, people often ignore these important communications. This paper explores the links between warning dialog design and user understanding of, motivation to respond to, and actual response to computer security warnings. We measured these variables through a 733-participant online study that tested a set of four existing computer security warnings and two redesigned versions of each across low- and high-risk conditions. In some cases our redesigned warnings significantly increased participants' understanding and motivation to take the safest action; however, we were not able to show that participants' responses were differentiated between low and high risk conditions. We also observed that motivation seemed to be a more important predictor of taking the safest action than understanding. However, other factors that may contribute to this behavior warrant further investigation.

Keywords: Security warning dialog, usable security.

1 Introduction

Warnings are communications designed to protect people from harm [1]. These harms may be immediate, as in the case when road signs warn about sharp turns, or they may be in the future, as in the case of health notices on cigarette boxes. In the case of computer security warnings, the harms arise from immediate and future threats to personal information (e.g., financial data) or property (e.g., computers). However, despite this threat of harm, people often do not read or understand computer security warnings [2-4] and frequently fail to heed them [5], even when the situation is hazardous. There is a lack of empirical evidence about the factors that influence response to computer warnings [6].

This paper uses the results of a 733-participant online study based on a set of existing and redesigned warnings to examine the links between warning design, user understanding of risk, motivation to respond to the risk, and decision to take the least risky action. In this paper, we focus on computer security dialogs, a subset of security warnings, which are warnings that offer users a choice of at least two options.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 18–35, 2011.

1.1 Warnings Research

In the warnings literature, response to a warning is often evaluated in terms of 'compliance' – performing an action when instructed to do so [7]. Much of the prior research on computer security compliance behaviors focused on phishing attacks or web browser certificates. In one study, over two-thirds of participants, in a laboratory setting, dismissed anti-phishing warnings because of the websites' look-and-feel and participants' incorrect mental models [8]. Similarly, in an online survey, between 30% and 60% of participants said they would ignore each of the tested browser warnings and continue to a potentially dangerous website. In a subsequent laboratory study, redesigned versions of these warnings achieved greater compliance, but, even in the best case, 45% of participants still ignored the warning when it interfered with their primary tasks [4]. In another laboratory study, about half of the participants ignored a full-page warning presented before an authentication page to an online banking website [5]. Although this behavior may be considered rational from an economic perspective, the problem of how to design effective security communications that do not burden users still remains [9].

Previous research shows high levels of warning non-compliance, even after warning redesign, providing only limited insights into the reasons for non-compliance. People might fail to heed warnings for a variety of reasons, including lack of trust in the warnings, lack of awareness of the risks [2], lack of understanding of the warnings [10], and lack of motivation to comply (perhaps because the required effort is larger than the benefit [9]). Potential consequences for lay users include the possibility of becoming a victim of phishing and other types of scams, of downloading a virus and losing information, of disclosing private and sensitive information, or of being exposed to other harmful threats. This study goes beyond prior research to examine two possible causes of non-compliance: lack of understanding and lack of motivation.

Previous work suggests that lack of understanding may contribute to non-compliance. Egelman et al. observed that some participants who encountered web browser phishing warnings after receiving a phishing email still believed the emails were legitimate. The authors describe a "cognitive dissonance" between the spoofed website and the email leading to it [3]. Motieé et al. reported that 77% of all participants to a laboratory study did not understand the purpose of security warnings and consented to a fake security prompt [11]; in same study, 22% of participants with high level of computer expertise did the same.

There are also qualitative theoretical models that apply to how users interact with computer security warnings. The Communication-Human Information Processing (C-HIP) model [12] describes the human processes involved in the internalization of a warning. In the model, a warning travels from a source through a channel to a receiver. The model focuses on a set of sequential stages–attention switch, attention maintenance, comprehension/memory, attitudes/beliefes, and motivation–through which a receiver processes the warning, resulting in a behavior. The Human-In-The-Loop security framework, based on the C-HIP model [13], can be used to systematically identify security issues created by users who fail to properly carry out security-critical tasks. This framework predicts errors in cases where users do not know how to comply or are unmotivated or incapable of complying with warnings

[13]. This study was designed to examine parts of this framework; specifically, it investigates the relationship between understanding, motivation, and user response.

1.2 Safe Response

While some previous work talks about warning 'compliance', we use the term 'safe response' instead. Safe response is an objective measure, that is defined as taking the least risky option provided by a computer security warning dialog. For example, one of the warnings used in this study warns about the possibility that an email attachment may infect a user's computer with a virus. The safe response would be not to open the email attachment, as this is the only response that would present no risk to the user. Any other response, such as opening or saving the attachment, would present some level of risk.

Safe response differs from compliance, which is a concept borrowed from research into physical, non-interactive warnings [7]. In the case of security warning dialogs, we feel that safe response is a clearer metric. In computer systems, there are many situations that may be more or less safe, depending on a context known only to a user. Well-designed security warnings tend to address such situations, as any hazards that could be addressed without contextual knowledge should have been blocked without user intervention. A good security warning will assist the user in using her contextual knowledge to make an informed choice between two or more options, one of which is the least risky option, or the 'safe response.'

High levels of safe response are not always necessary. There is a trade-off between usability and level of risk that is based on the specific context. Always making the least risky choice would allow for a completely safe system but would reduce functionality. A warning is useful if it helps a user to use her knowledge of the context to make an informed decision that balances risk and usability. For example, in the attachment warning outlined above, the 'safe response' would be to not open the attachment. However, within a given context the user should consider factors exogenous to the system, determine how risky the context is, and decide if she should open the attachment. If the user is expecting a file, knows the sender, and can tell from the warning text that this is the file she was expecting, then she finds herself in a low-risk context. In this particular context, the safe response is not necessary and she should open the attachment.

We analyze safe response as being a desirable response in high-risk contexts, under the assumption that users should protect themselves against the high risk of a potential threat, and as being an undesirable response in low risk contexts, under the assumption that it is unnecessary for users to block functionality in these situations. Sunshine et al. took a similar approach in their evaluation of user response to web browser certificate warnings on an online banking login page (high risk) and a university library website (low risk) [4].

2 Methodology

We performed an online survey (n=733) to test the effects of warning design on user understanding, motivation, and safe response. Our study used a 3 x 2 design, with three warning design conditions (E: existing warnings, G: redesigned based on warning design guidelines, and M: redesigned based on our previous work on mental models) and two scenario-based context conditions (S_1: low security priming and S_2: high security priming) for a total of six conditions.

2.1 Warning Design Conditions

We tested five existing warnings from commercially available software, but report on only the four that are security dialogs. The four warnings, referred to as the *Existing set* (E, see Figure 5 in the Appendix), alerted users about problems encrypting an email (W1), a program trying to access the user's address book (W2), an email attachment (W3), and an unknown certificate (W4).

Table 1. Guidelines used to redesign warnings

Guideline	Examples
1. Follow a visually consistent layout	Use one icon; do not use a close button; use command links for options; use a primary text to explain the risk; describe the consequences of each option below each button.
2. Comprehensively describe the risk	Describe the risk; describe consequences of not complying; provide instructions on how to avoid the risk.
3. Be concise, accurate and encouraging	Be brief; avoid technical jargon; provide specific names, locations and values for the objects involved in the risk; do not use strong terms (e.g., abort, kill, fatal)
4. Offer meaningful options	Provide enough information to allow the user to make a decision; option labels should be answers to explicit question asked to the user; if only one option is available, do not show the warning; the safest option should be the default.
5.0020Present relevant contextual and auditing information	If the warning was triggered by a known application, describe the application; identify agents involved in the communication by name; if user's information is about to be exposed to risk, describe what information and how it will be exposed.

We created a second set of warnings, referred to as the *Guideline-based set* (G, see Figure 5 in Appendix). Each of the warnings in the E set were redesigned by three HCI Master's students who each had at least one year of HCI coursework as well as previous design experience. We asked the students to redesign the existing warnings by following design guidelines that we compiled from the literature [3, 12-19]. A brief summary of these guidelines is shown in Table 1. We did not provide the designers with any other information about our study.

Similarly, we created a third set of warnings, referred to as the *Mental-model-based set* (M, see Figure 4 in Appendix). To create this set we redesigned each warning in the E set based on previous work on mental models of computer security warnings. In our previous work we found differences in the way experts and non-experts respond to these warnings [20]. We tried to design this set of warnings to include information that experts tend to seek out when responding to a warning, such as the results of analyses by anti-virus software. We also applied many of the guidelines used by the HCI students to create set G.

2.2 Contextual Scenarios

Users view security warning dialogs within a specific contextual situation,and make a decision based on that situation. To imitate this context in our online survey, we wrote a *Scenario 1* (S_1) and a *Scenario 2* (S_2) for each warning. Each user who saw a particular warning was presented with a scenario along with that warning. S_1 included low security-priming scenarios with activities that most people would not normally associate with a security threat; whereas, S_2 included activities that involved sensitive or confidential information, or had characteristics of common security attacks. As warnings must consistently be useful in both low- and high-threat contexts we chose to include both low and high security-priming categories to ensure that our results were consistent across scenarios that presented different threat levels. Table 2 contains all scenarios. We incorporated feedback from security experts when creating the scenarios and strove to ensure that scenarios were of similar readability and length.

Table 2. Scenarios created for the study

	Low risk / Scenario 1 (S1)	Scenario 2 (S2) / High risk
W1: Encryption warning	Imagine that you are sending a birthday greeting to your friend Rob by email. You click on the 'Send' button and the warning below appears on your screen.	Imagine that you are sending important financial information to your boss by email. Your boss warned you that it is important to keep this information confidential. You click on the 'Send' button and the warning below appears on your screen.
W2: Address book warning	Imagine that you are trying to connect your PDA or smartphone to your computer to synchronize your email. You plug the device into your computer, and the warning below appears on your screen.	Imagine that you are reading your email. You open a message from your friend Rob, and the message invites you to try out a new social network your friend is using. You click on the invitation, and the warning below appears on your screen.
W3: Attachment warning	Imagine you are reading your email. You open an email from a friend, who says that he is sending you a book he thinks you would find interesting. You double-click on the attachment, and the warning below appears on your screen.	Imagine you are reading your email. You open an email that seems to be from one of your friends, but the email does not contain any text, only a document attached. You double-click on the attachment and the warning below appears on your screen.
W4: Certificate warning	Imagine that you want to buy a gift for a very good friend, but you don't have time to go to a store. You look for a site on the Web, and after searching for a few minutes you find a website that seems to be OK. You click on the link to the website, and the warning below appears on your screen.	Imagine that you need to pay a bill, and you are out of checks. A friend suggests you try paying the bill from your bank's website. You have seen your bank statements online before, but you don't know how to pay bills online. You remember you recently received an email from your bank. You open the email, click on a link to enter the bank's website, and the following warning appears.

2.3 High and Low Risk Conditions

Each warning, in combination with each scenario, presented the user with either a high or low level of risk. Throughout this paper, we refer to the level of risk that the participant faced when presented with a specific warning and contextual scenario combination as either Low Risk (LR) or High Risk (HR). Based on our definition of safe response, when warnings are successful, participants in LR conditions should choose not to take the safe response because the safe response requires them to sacrifice functionality. However, participants in HR conditions should choose the safe response because they should prioritize safety over functionality in risky situations.

We had two low-risk conditions: the encryption and address book warnings with S_1 scenarios. In both cases the risk is minimal and taking the least risky action would prevent the user from completing her primary task. We had six high-risk conditions: all four warnings with S_2 scenarios, and the attachment and certificate warning with S_1 scenarios.[1] In these cases, the level of risk warranted taking the safe response.

A well-designed security dialog should allow participants to differentiate between low- and high-risk conditions. It should create a higher rate of motivation and safe response for high-risk conditions than for low-risk conditions. If the warnings in our study were well designed we would expect to see warnings with the same level of risk in S_1 and S_2 (attachment and certificate warnings) to have similar rates of motivation and safe response. We would also expect to see warnings with low risk in S_1 and high risk in S_2 (encryption and address book warnings) to have higher levels of safe response and motivation in S_2.

Table 3. Number of participants in each condition

	W1		W2		W3		W4	
Scenario	S_1	S_2	S_1	S_2	S_1	S_2	S_1	S_2
E	145	124	114	145	125	114	106	125
G	119	106	124	119	145	124	114	145
M	125	114	106	125	119	106	124	119

2.4 Survey Design and Participant Recruiting

Our survey consisted of 69 questions divided into seven sections, starting and ending with demographic questions. Each of the remaining five sections included a randomly selected image of a warning, a randomly selected corresponding scenario (S_1 or S_2), and a set of questions about each warning.

We recruited participants using Amazon's Mechanical Turk service [20], paying each participant who completed the study 50 cents. We required participants to be computer users, over 18 years old, English-speaking and residents of the United States. Participants took an average of 10 min 47 sec to answer the survey (σ = 7 min 9 sec). We discarded 3 responses that took less than 10 seconds. We were left with 733 respondents, about 62% of whom were females and four-fifths of whom were Caucasian. The number of participants in each condition is summarized in Table 3. Participants ranged in age from 18 to 75, with a mean age of 32.9 (σ = 11.58). We

[1] The content of the attachment and certificate warnings (see Appendix) was suspicious enough to suggest a high-risk situation, even in S_1.

also collected information about usage of operating systems, browsers, and email clients to test any correlation with our dependent variables. As described later, we found no consistent relationship between demographics and dependent variables.

We also asked two questions to probe participants' level of technical expertise: whether they had ever taken or taught a course on computer security, and whether they knew any computer languages. If they answered the latter affirmatively, we asked which languages they knew. Participants who answered only HTML were not considered as having programming expertise. We found no significant correlation between affirmative answers and any studied variables, so we excluded these questions from our analyses.

Table 4. Questions asked to participants per warning, and the corresponding measured variable

Dependent variable	Question	Types of answers	Explanation
Under-standing	*What do you think is/are the problem(s)?*	11 common problems plus an *Other* open text field	If participants answered at least one of the correct answers and none of the incorrect answers (based on authors' knowledge and interviews with security experts [20]), understanding was recorded as 1, otherwise as 0.
Motivation	*The problem described by this warning is very important.*	5-point Likert response, from *Strongly disagree* to *Strongly agree*	If participants answered *Agree* or *Strongly agree*, motivation was recorded as 1, otherwise as 0.
Safe response	*What would you do in this situation?*	As many clickable options as the warning offered, plus *Ignore this warning* and *Take another action*	If participants answered at least one action considered safe by experts and none of the actions considered unsafe by experts, safe response was recorded as 1, otherwise as 0.

2.5 Hypotheses

To develop our hypotheses, we defined three dependent variables: understanding, motivation and safe response. These variables are described in Table 4. We also defined low- and high-risk conditions consisting of combinations of warnings and scenarios, as given below:

Low-risk condition: W1 with S_1, W2 with S_1.

High-risk conditions: W1 with S_2, W2 with S_2, W3 with S_1 or S_2, and W4 with S_1 or S_2.

We hypothesized that understanding would be higher for all conditions in the redesigned warnings than in the existing set. For motivation and safe response we hypothesized that they would be significantly higher in the redesigned warnings for participants in the high-risk conditions but would not be significantly higher for participants in the low-risk condition. We also hypothesized that understanding and motivation would be found to drive safe response. Our hypotheses are enumerated below:

H_1: For all warnings and scenarios, understanding will be significantly higher in the guidelines-based (G) and mental-model-based (M) sets than in the existing set (E).

H_2: For all low-risk scenarios, motivation and safe response will not be significantly higher in the redesigned sets (G and M) than in the existing set (E).

H_3: For all high-risk scenarios, motivation and safe response will be significantly higher in the redesigned sets (G and M) than in the existing set (E).

H_4: Understanding and motivation will be significant predictors of safe response across all warning sets and scenarios, controlling for demographic factors.

3 Analysis

Based on an analysis of the four warnings we found that understanding and motivation were strongly correlated with safe response. However, we were not able to conclude that users could differentiate between low-risk and high-risk conditions, and we did not see a significant increase in motivation and safe response for W1 and W2 in either the high- or low-risk conditions. However, we did find improvements in motivation and safe response for W3 and W4, the two warnings that were only presented in high-risk conditions.

We analyzed our results separately for each warning using logistic regression. Logistic regression is similar to linear regression except that it is used to analyze data with a binary dependent variable. Factors with significant p-values are significant predictors of the dependent variable, controlling for all other factors (see Tables 4 and 5 in Appendix). We used a significance level of $\alpha = .05$ for all analyses.

3.1 Understanding

In general, our redesigned sets of warnings (G and M) failed to increase understanding over existing warnings. We observed significant increases in understanding in only 3 out of 16 conditions, and in two cases related to W2 we observed significant decreases in understanding. Figure 1 shows our results for understanding. Statistical data are given in Table 5 in the Appendix.

We expected to see increased levels of understanding for the G and M sets versus the E set (H_1). While this occurred in a few conditions, understanding did not increase in the majority of cases (see Table 5 in Appendix). Because understanding increased in more conditions in which participants were shown S_1 than S_2, we tested the possibility that participants spent less time on the scenarios by comparing the mean time that participants took to answer each warning section. However, we found no significant differences between times for the two sets of scenarios.

In the S_1 scenario for the address book warning (W2), the understanding rate was significantly lower for the G and M sets than in the E set. To help explain this lower level of understanding we looked at the specific problems that users thought the warning presented. We found that a higher percentage of respondents believed that the warning was related to a website in the G and M sets than in the E set, which was

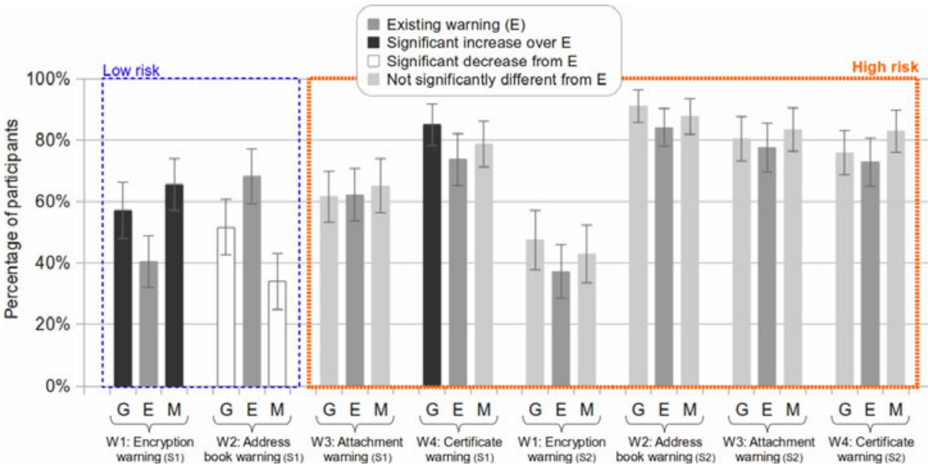


Fig. 1. Percentage of participants who showed understanding of the problem that triggered the studied warnings, in the low- and high-risk conditions. G, E, and M correspond to the different sets of warnings. The top bars represent confidence intervals at the 95% level.

a "wrong" answer. The misunderstanding was potentially due to a reference to ABC.exe (the program accessing the computer) that only appeared in the redesigned warnings. We speculate that respondents may have mistaken ABC.exe for a website. We mandated in our guidelines that a program prompting a warning should be identified to users, to help them better decide how to respond, but the implementation of this recommendation could have resulted in confusion.

The redesigned warnings (G and M) were also less likely to prompt two 'right' answers than the existing (E) warning. For the G and M versions of the address book warning in the S_1 scenario, participants were less likely to respond that they did not trust the software being run or that there was no problem than when shown the E version of the warning. Participants may not have considered ABC.exe to be software, or perhaps they considered the redesigned warnings more threatening than the existing warning. Additional testing is necessary to determine which aspects of the warnings lead to misunderstanding.

These results provide very limited, if any, support for H_1. It should be noted, however, that many warning-scenario combinations had a high initial level of understanding, from which it may be difficult to introduce improvements.

3.2 Motivation

Our redesigned warning sets (G and M) had some success at increasing levels of motivation in the high-risk condition for W3 and W4, but did not show evidence of allowing participants to differentiate between low- and high-risk conditions. Figure 2 shows our results for motivation. Statistical data are given in Table 6 in the Appendix.

If the redesigned warnings allowed participants to differentiate between high- and low-risk contexts and respond appropriately, there would be no change in motivation levels between G/M and E in the low-risk condition, but there would be an increase in motivation levels for the redesigned warnings in the high-risk condition. We were not able to conclude that the redesigned warnings allowed users to differentiate between low- and high-risk contexts. For the encryption warning and address book warning (W1 and W2), which were shown in both high- and low-risk contexts, there was no significant improvement in motivation in the majority of cases in either context.

In the low-risk context we expected motivation not to be significantly higher for the redesigned warnings (G and M) than the existing warnings (E). This held for three out of four cases, providing support for H_2. However, for these results to be meaningful, we needed to see a corresponding increase in motivation for these same warnings (W1 and W2) in a high-risk context, proving that participants could differentiate between the levels of risk with the redesigned warning set and respond appropriately. However, we found that in all four high-risk cases for W1 and W2 there was no significant difference between the E set and each of the G and M sets for motivation. This indicates that the lack of improvement in the low risk case may have represented a lack of improvement overall, rather than participants' abilities to differentiate between risk levels. Thus, while these results support H_2, they are inconclusive.

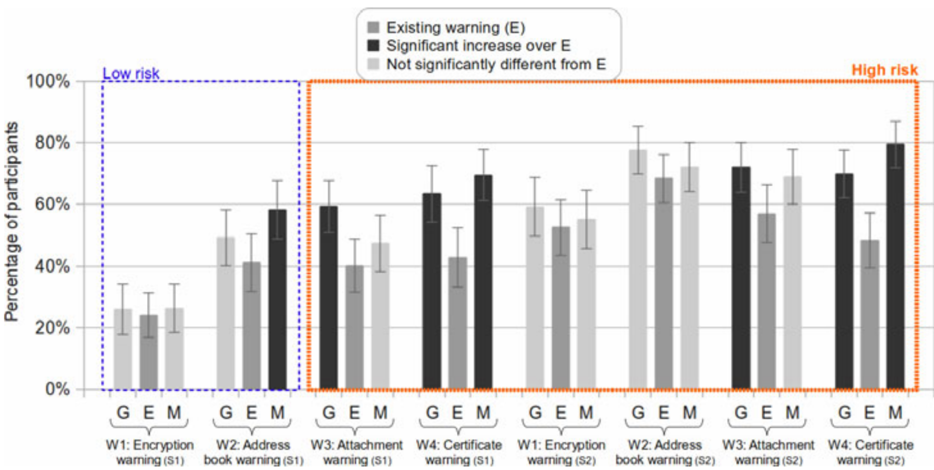


Fig. 2. Motivation, as measured by the percentage of participants who agree or strongly agree that the problem described by the warning is very important, in the low- and high-risk conditions. G, E, and M correspond to the different sets of warnings. The top bars represent confidence intervals at the 95% level.

Although there was no evidence that the redesigned warnings allowed participants to differentiate between low- and high-risk contexts, we did find some evidence that the redesigns improved motivation in the high-risk context (H_3). For the attachment and certificate warnings (W3 and W4), which were only shown in high-risk contexts, we found that the redesigned warnings significantly increased motivation in all but one case. As previously described, we expected to see similar results for W1 and W2 in the high-risk context, but did not see any significant differences between G/M and E for W1 and W2.

3.3 Safe Response

We found that the redesigned warnings were successful at increasing safe response in the majority of the high risk conditions. However, as was the case with motivation, we were not able to conclude that the redesigned warnings allowed participants to differentiate between high- and low-risk conditions and respond appropriately. Figure 3 shows our results for safe response. Statistical data are given in Table 6 in the Appendix.

As described previously, safe response measures the proportion of participants who pick the option that presents the least risk. We expected participants' rates of safe response to significantly increase for the high-risk conditions for our redesigned warnings and to remain the same for the low-risk conditions. In the low-risk conditions the redesigned warnings should not push participants to pick a safe response that would prevent them from completing the desired task. For the two warnings that we presented in both the high- and low-risk conditions, W1 and W2, we found that, as expected, in three out of four cases, the level of safe response was not higher for the G and M sets than for the E set. However, for these two warnings we also found that, in three out of four cases, the level of safe response did not increase in the high-risk condition for G and M compared to E, indicating that the lack of improvement in the low-risk condition may have been due to an overall lack of improvement rather than participants' ability to differentiate between risk levels. So, although we found some evidence for H_3, our overall results for safe response for warnings W1 and W2 were inconclusive.

We did, however, find a significant increase in safe response levels for the redesigned warnings (G and M) over the existing set (E) for the two warnings that were presented in only the high risk condition, W3 and W4. For these warnings, rates of safe response significantly increased in seven out of eight cases.[2] This result provides some support for H_3.

3.4 Correlation between Variables

We hypothesized that understanding and motivation would be predictors of safe response (H_4). We found significant correlation between safe response and understanding, motivation, and other variables (see Table 6 in Appendix), supporting H_4. The higher logistic regression coefficients show that safe response is strongly tied to motivation and also linked, although slightly less strongly, to understanding. Although these results do not prove that understanding and motivation drive safe response, they provide some indication that the variables are strongly related.

Motivation and understanding were significantly correlated with each other for all warnings. Motivation was also significantly correlated with safe response for all four warnings for all warning sets. Understanding was also significantly correlated with safe response for all except the encryption warning (W1). Based on the regression coefficients, motivation was more strongly correlated with safe response for all of the warnings in which both factors were significant, except for the address book warning (W2).

---

[2] We performed a qualitative analysis of participants' open comments at the end of each warning to test the possibility that these higher levels of safe response were due to the novelty of redesigned warnings. We found no evidence of such behavior.

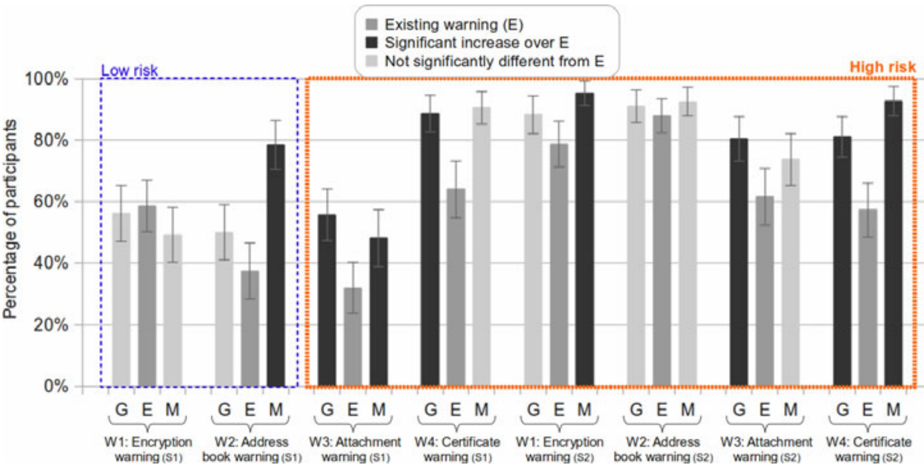


Fig. 3. Percentage of participants who took the safest option, in the low- and high-risk conditions. G, E, and M correspond to the different sets of warnings. The top bars represent confidence intervals at the 95% level.

Outside of motivation and understanding, we also found interactions between age and being a user of Microsoft Internet Explorer for the address book (W2) and the certificate (W4) warnings. This was expected, as these users have likely encountered these warnings before. In the address book warning, users of Internet Explorer were more likely to pick the safest response, while in the certificate warning (W4), the opposite relation held.

4 Discussion

One of the primary goals of this study was to show differentiated results for low- and high-risk conditions to demonstrate that our redesigned warnings improved participants' abilities to make appropriate security choices in each of the conditions. However, our results did not show differentiated motivation and safe response improvements for the low- and high-risk conditions. For both of the warnings that were presented in low- and high-risk conditions (W1 and W2) we found that in the majority of cases motivation and safe response did not significantly increase for the redesigned warnings in both conditions. It is likely that the redesigned warnings were not more effective than existing warnings and were not able to increase motivation or safe response in either case. It is also possible that the high security-priming scenarios that were used to prompt the high-risk condition were poorly designed and did not prompt a high-risk response. However, this is less likely as 3 out of 8 had significantly higher levels of motivation and safe response for the high-risk condition. Further research is needed to better determine how users respond to high- and low-risk conditions and how to consistently design better security warning dialogs.

One of our redesigned warnings, the M version of the address book warning (W2), turned out to be particularly ineffective. It decreased participants' understanding, increased user motivation and safe response in the low-risk condition, and did not increase motivation or safe response in the high-risk condition. One potential explanation for this unexpected behavior is the amount of information that version contained: the existing version had 44 words and 4 options, and the guidelines-based version had 40 words and 3 options, while the mental-model-based version had 163 words and 6 options. The extra text included the results of an anti-virus scan, and an explanation of the consequences for each option. The large amount of information may have undermined participants' abilities to understand (or motivation to read) the redesigned warning, or some element of the added text might have confused them.

Although our redesigned warnings appear not to help participants differentiate between high- and low-risk conditions, we were able to demonstrate that it is possible to use a relatively simple redesign process to improve some security warning dialogs for high-risk conditions. Beyond the importance of testing whether participants could differentiate between high- and low-risk conditions, it was also important to show that our results were applicable across different types of contextual scenarios. To do so, we presented participants with low and high security-priming contexts (S_1 and S_2). Further work is necessary to determine which aspects of the redesigns contributed to the successful increases in motivation and safe response and which aspects were not successful at increasing understanding, motivation and safe response.

4.1 Limitations

Our study had a variety of limitations, some of which we hope to improve upon in future work. First, the study is based on self-reported survey data, and as such it may not reflect what users would do when confronted with warnings during their regular computer use. Also, literature suggests that habituation should be considered when studying warnings [12]. To the best of the authors' knowledge, repeated, long-term exposure to computer warnings has not been studied, in part because of the difficulties in setting up adequate experimental designs. However, a deeper look at the answers of our participants show that only a small proportion of them reported that they ignored our warnings, either because they had seen them before or for other reasons. If our participants had been habituated to our set of existing warnings, we would expect to have seen a higher number of people ignoring them. Another factor that might have affected participants' response is the novelty of redesigned warnings. Although we found no evidence in this direction, this remains a limitation of our study.

Another confounding factor might be the possible learning process that takes place after repeated exposures to the same set of questions with different warnings. A technical limitation of the software we used to implement the survey[3] prevented us from tracking the random order in which participants saw our warnings. Although randomization might counter-balance learning effects, we acknowledge that this does not necessarily cancel out the effects. One improvement to the experimental design would be to show a single warning to each participant. We decided to show five warnings instead of one to reduce the number of participants needed for the study.

Our redesigned sets utilized different layouts of options, longer and more descriptive texts for each option, information about context, and the results of analysis

---

[3] SurveyGizmo, available at `http://www.surveygizmo.com`

by other tools. However, our experimental design did not allow us to isolate the impact of each of these design changes. In future work we expect to better isolate specific factors.

4.2 Conclusion

By comparing existing computer security warnings with two sets of warnings that we created, we explored relationships between the design of the warning, understanding of the problem underlying a warning, the belief that the problem is important (motivation), the tendency to pick the safest option (safe response), and demographic factors. We found that design changes can lead to improvements in understanding, motivation, and tendency to pick the safest option in some cases, but further work is needed to isolate the impact of various design factors. However, we were unable to help participants differentiate between the appropriate option in high- and low-risk conditions. We also found that although understanding and motivation are strongly tied to each other, motivation is a slightly more important factor than understanding when it comes to increasing safe response to warnings.

Warning designers should keep in mind that both the level of importance that users attribute to a warning and the understanding of the problem underlying a warning contribute to user response. To be successful, warnings should both motivate a user to respond, and help users understand the risk, in that order. Future work should look at exactly how much each of these factors, and other factors, contribute to increasing safe response to warnings.

Acknowledgements. This research was funded in part by NSF grant CNS0831428. The last author wishes to thank the ARCS Foundation for their support.

References

[1] Wogalter, M.S.: Purposes and scope of warnings. In: Wogalter, M.S. (ed.) Handbook of Warnings. Human Factors and Ergonomics, 1st edn., pp. 3–9. Lawrence Erlbaum Associates, Mahwah (2006)
[2] Downs, J.S., Holbrook, M.B., Cranor, L.F.: Decision strategies and susceptibility to phishing. In: Cranor, L.F. (ed.) Proceedings of the 2nd Symposium on Usable Privacy and Security (SOUPS). ACM International Conference Proceeding Series, vol. 149, pp. 79–90. ACM, New York (2006)
[3] Egelman, S., Cranor, L.F., Hong, J.I.: You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In: Czerwinski, M., Lund, A.M., Tan, D.S. (eds.) Proceedings of the 2008 Conference on Human Factors in Computing Systems (CHI), pp. 1065–1074. ACM, New York (2008)
[4] Sunshine, J., Egelman, S., Almuhimedi, H., Atri, N., Cranor, L.F.: Crying wolf: An empirical study of ssl warning effectiveness. In: Proceedings of the 18th Usenix Security Symposium (August 2009)
[5] Schechter, S.E., Dhamija, R., Ozment, A., Fischer, I.: The emperor's new security indicators. In: SP 2007: Proceedings of the 2007 IEEE Symposium on Security and Privacy, pp. 51–65. IEEE Computer Society, Washington, DC (2007)

[6] Camp, L.J.: Mental models of privacy and security. IEEE Technology and Society Magazine 28(3), 37–46 (2009)
[7] Meyer, J.: Responses to dynamic warnings. In: Wogalter, M.S. (ed.) Handbook of Warnings. Human Factors and Ergonomics, 1st edn., pp. 221–229. Lawrence Erlbaum Associates, Mahwah (2006)
[8] Wu, M., Miller, R.C., Garfinkel, S.L.: Do security toolbars actually prevent phishing attacks? In: Grinter, R.E., Rodden, T., Aoki, P.M., Cutrell, E., Jeffries, R., Olson, G.M. (eds.) Proceedings of the Conference on Human Factors in Computing Systems (CHI), pp. 601–610. ACM, New York (2006)
[9] Herley, C.: So long, and no thanks for the externalities: the rational rejection of security advice by users. In: Proceedings of the 2009 Workshop on New Security Paradigms Workshop, NSPW 2009, pp. 133–144. ACM, New York (2009)
[10] Downs, J.S., Holbrook, M.B., Cranor, L.F.: Behavioral response to phishing risk. In: Cranor, L.F. (ed.) Proceedings of the Anti-Phishing Working Groups 2nd Annual eCrime Researchers Summit 2007. ACM International Conference Proceeding Series, vol. 269, pp. 37–44. ACM, New York (2007)
[11] Motieé, S., Hawkey, K., Beznosov, K.: Do windows users follow the principle of least privilege?: investigating user account control practices. In: Proceedings of the Sixth Symposium on Usable Privacy and Security, New York, NY, USA, pp. 1–13 (2010)
[12] Wogalter, M.S.: Communication-human information processing model. In: Wogalter, M.S. (ed.) Handbook of Warnings. Human Factors and Ergonomics, 1st edn., pp. 51–61. Lawrence Erlbaum Associates, Mahwah (2006)
[13] Cranor, L.F.: A framework for reasoning about the human in the loop. In: Churchill, E.F., Dhamija, R. (eds.) Usability, Psychology, and Security, USENIX Association (April 2008)
[14] Apple Inc.: Apple human interface guidelines (2010), Online document available at `http://developer.apple.com`
[15] Benson, C., Elman, A., Nickell, S., Robertson, C.Z.: Gnome human interface guidelines 2.2.1 (2010), Online document available at `http://library.gnome.org` (last visit on April 08, 2010)
[16] Microsoft Corporation: Windows user experience interaction guidelines (2010), Online document available at `http://msdn.microsoft.com` (last visit on April 08, 2010)
[17] Egelman, S.: Trust me: Design Patterns for Constructing Trustworthy Trust Indicators. PhD thesis, School of Computer Science, Carnegie Mellon University (2009), Available as technical Report CMU-ISR-09-110
[18] Nodder, C.: Users and trust: a microsoft case study. In: Cranor, L.F., Garfinkel, S.L. (eds.) Security and Usability: Designing Secure Systems that People Can Use. Theory in Practice, 1st edn., pp. 589–606. O'Reilly Media, Inc., Sebastopol (2005)
[19] Ross, B.: Firefox and the worry-free web. In: Cranor, L.F., Garfinkel, S.L. (eds.) Security and Usability: Designing Secure Systems that People Can Use. Theory in Practice, 1st edn., pp. 577–587. O'Reilly Media, Inc., Sebastopol (2005)
[20] Bravo-Lillo, C., Cranor, L., Downs, J., Komanduri, S.: Bridging the gap in computer security warnings: a mental model approach. IEEE Security and Privacy Magazine (to appear, 2011)
[21] Ross, J., Irani, L., Silberman, M.S., Zaldivar, A., Tomlinson, B.: Who are the crowdworkers?: shifting demographics in mechanical turk. In: CHI EA 2010: Proceedings of the 28th of the International Conference Extended Abstracts on Human Factors in Computing Systems, pp. 2863–2872. ACM, New York (2010)

Appendix

Table 5. Comparison of percentage of participants that showed understanding (top), motivation (middle) and safe response (bottom) between warning sets. Black cells show significant increases over existing set, dark grey show significant decreases from existing set, and light gray cells show non-significant differences from existing set. c is coefficient, SE is standard error, z is z-value, and p is p-value.

	Scenario 1 (S1)		Scenario 2 (S2)	
W1: Encryption warning	E (41%) < G (57%) c=0.668; SE=0.260; z=2.569; p=.010	E (41%) < M (66%) c=1.025; SE=0.260; z=3.945; p<.001	E (37%) ~ G (48%) c=0.423; SE=0.274; z=1.542; p=.123	E (37%) ~ M (43%) c=0.238; SE=0.273; z=0.871; p=.384
W2: Address book warning	E (68%) > G (52%) c=-0.696; SE=0.278; z=-2.508; p=.012	E (68%) > M (34%) c=-1.428; SE=0.294; z=-4.859; p<.001	E (84%) ~ G (91%) c=0.648; SE=0.408; z=1.590; p=.112	E (84%) ~ M (88%) c=0.291; SE=0.364; z=0.799; p=.424
W3: Attachment warning	E (62%) ~ G (62%) c=-0.027; SE=0.258; z=-0.105; p=.916	E (62%) ~ M (65%) c=0.125; SE=0.273; z=-0.458; p=.647	E (78%) ~ G (81%) c=0.178; SE=0.328; z=0.541; p=.588	E (78%) ~ M (83%) c=0.380; SE=0.352; z=1.079; p=.280
W4: Certificate warning	E (74%) < G (85%) c=0.703; SE=0.352; z=2.00; p=.046	E (74%) ~ M (79%) c=0.279; SE=0.318; z=0.878; p=.380	E (73%) ~ G (76%) c=0.157; SE=0.288; z=0.547; p=.585	E (73%) ~ M (83%) c=0.596; SE=0.324; z=1.840; p=.066

	Scenario 1 (S1)		Scenario 2 (S2)	
W1: Encryption warning	E (24%) ~ G (26%) c=0.098; SE=0.296; z=0.330; p=.741	E (24%) ~ M (26%) c=0.115; SE=0.289; z=0.399; p=.690	E (53%) ~ G (59%) c=0.271; SE=0.272; z=0.996; p=.319	E (53%) ~ M (55%) c=0.105; SE=0.268; z=0.390; p=.696
W2: Address book warning	E (41%) ~ G (49%) c=0.325; SE=0.296; z=1.207; p=.227	E (41%) < M (58%) c=0.692; SE=0.280; z=2.470; p=.014	E (68%) ~ G (78%) c=0.474; SE=0.294; z=1.613; p=.107	E (68%) ~ M (72%) c=0.178; SE=0.275; z=0.647; p=.518
W3: Attachment warning	E (40%) < G (59%) c=0.779; SE=0.256; z=3.049; p=.002	E (40%) ~ M (47%) c=0.291; SE=0.264; z=1.102; p=.270	E (57%) < G (72%) c=0.664; SE=0.283; z=2.344; p=.019	E (57%) ~ M (69%) c=0.515; SE=0.289; z=1.782; p=.075
W4: Certificate warning	E (43%) < G (64%) c=0.849; SE=0.283; z=3.00; p=.003	E (43%) < M (69%) c=1.117; SE=0.282; z=3.956; p<.001	E (48%) < G (70%) c=0.909; SE=0.262; z=3.472; p=.001	E (48%) < M (79%) c=1.419; SE=0.296; z=4.795; p<.001

	Scenario 1 (S1)		Scenario 2 (S2)	
W1: Encryption warning	E (59%) ~ G (56%) c=-0.098; SE=0.259; z=-0.378; p=0.7054	E (59%) ~ M (49%) c=-0.382; SE=0.253; z=-1.513; p=0.1303	E (79%) ~ G (88%) c=0.712; SE=0.381; z=1.870; p=.061	E (79%) < M (95%) c=1.702; SE=0.510; z=3.334; p=.001
W2: Address book warning	E (37%) < G (50%) c=0.516; SE=0.272; z=1.898; p=0.0576	E (37%) < M (79%) c=1.819; SE=0.313; z=5.819; p<5.9e-09	E (88%) ~ G (91%) c=0.333; SE=0.425; z=0.783; p=.434	E (88%) ~ M (93%) c=0.541; SE=0.437; z=1.237; p=.216
W3: Attachment warning	E (32%) < G (56%) c=0.982; SE=0.261; z=3.761; p<.0001	E (32%) < M (48%) c=0.684; SE=0.271; z=2.523; p=.012	E (62%) < G (81%) c=0.942; SE=0.306; z=3.081; p=.002	E (62%) ~ M (74%) c=0.559; SE=0.300; z=1.865; p=.062
W4: Certificate warning	E (64%) < G (89%) c=1.490; SE=0.369; z=4.040; p<.0001	E (64%) < M (91%) c=1.696; SE=0.377; z=4.495; p<.0001	E (57%) < G (81%) c=1.166; SE=0.288; z=4.053; p<.001	E (57%) < M (93%) c=2.268; SE=0.410; z=5.530; p<.001

Table 6. Logistic regression coefficients of interactions between variables (H_4), per warning. Dark cells show significant, positive values, and grey cells show significant negative values.

	W1: Encryption warning	W2: Address book warning	W3: Attachment warning	W4: Certificate warning
Understanding	c=0.2693; se=0.3743; z=0.720; p=0.4718	**c=1.7099; se=0.4317; z=3.961; p=7.46e-05**	**c=0.7567; se=0.1911; z=3.959; p=7.53e-05**	**c=0.4945; se=0.2277; z=2.172; p=0.0298**
Motivation	**c=0.9021; se=0.3670; z=2.458; p=0.0140**	**c=1.4442; se=0.4158; z=3.473; p=0.00051**	**c=1.6107; se=0.1751; z=9.195; p<2e-16**	**c=1.6113; se=0.2125; z=7.582; p=3.41e-14**
Gender	c=-0.4687; se=0.3420; z=-1.370; p=0.1706	c=-0.0371; se=0.4255; z=-0.087; p=0.9304	c=-0.2056; se=0.1799; z=-1.143; p=0.2532	c=-0.1445; se=0.2096; z=-0.690; p=0.4904
Age	c=-0.0045; se=0.0223; z=-0.204; p=0.8384	c=0.0724; se=0.0451; z=1.606; p=0.1083	**c=0.0347; se=0.0150; z=2.313; p=0.0207**	c=0.0003; se=0.0156; z=0.025; p=0.9802
Use of Internet Explorer	c=0.6684; se=1.0370; z=0.645; p=0.5192	**c=2.8604; se=1.4160; z=2.020; p=0.0433**	c=-0.2554; se=0.5716; z=-0.447; p=0.6549	**c=-1.7783; se=0.6596; z=-2.696; p=0.0070**
Use of Internet Explorer – Age	c=-0.0024; se=0.0296; z=-0.082; p=0.9347	c=-0.0837; se=0.0501; z=-1.670; p=0.0948	c=-0.0035; se=0.0177; z=-0.200; p=0.8414	**c=0.0536; se=0.0205; z=2.608; p=0.0091**

Fig. 4. Mental-model-based (M) set of warnings

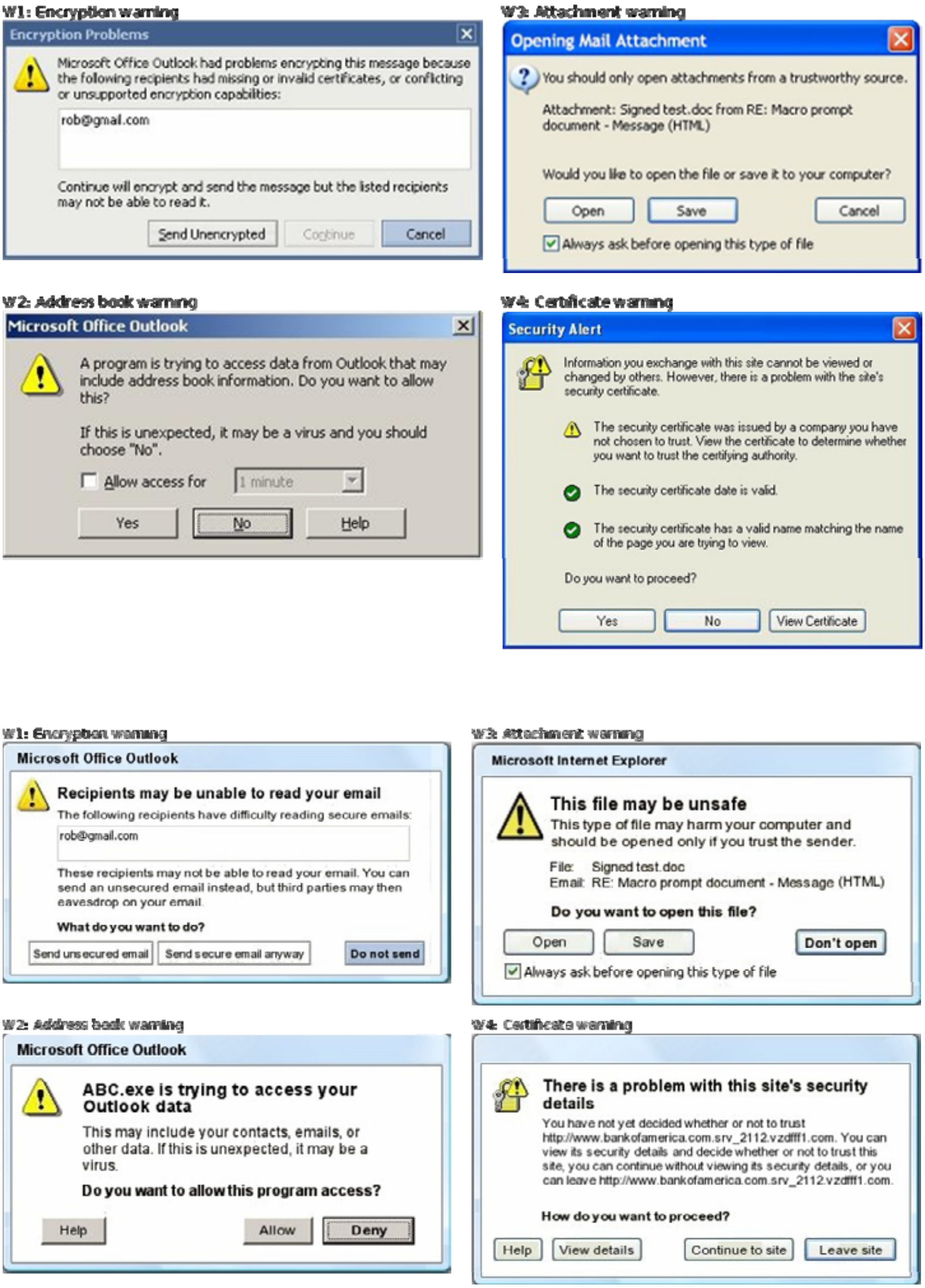


Fig. 5. Existing (E, top) set and Guidelines-based (G, bottom) sets of warnings

Usable Privacy and Security in Personal Health Records

Inma Carrión, Jose L. Fernández-Alemán, and Ambrosio Toval

Department of Informatics and Systems, University of Murcia, Spain
{mariainmaculada.carrion,aleman,atoval}@um.es

Abstract. PHRs (Personal Health Records) store individuals' personal health information. Access to this data is controlled by the patient, rather than by the health care provider. Companies such as Google and Microsoft are establishing a leadership position in this emerging market. In this context, the need for psychological acceptability in privacy and security protection mechanisms is essential. Any privacy and security mechanism must be acceptable from a usability perspective. This paper presents a study of the privacy policies of 22 free web-based PHRs. Security and privacy characteristics have been extracted according to the ISO/TS 13606-4 standard. In general, quite a good level was observed in the characteristics analyzed. Nevertheless, some improvements could be made to current PHR privacy policies to enhance the management of other users' data, the notification of changes to the privacy policy to users and the audit of accesses to users' PHRs.

Keywords: Usable privacy, usable security, PHRs, healthcare.

1 Introduction

In recent years, governments around the world have shown an increasing interest in computerizing health-care records [1]. The growing use of Web 2.0 technologies signifies that patients can now access their own health information via tools such as Personal Health Records (PHRs). The Markle Foundation defines a PHR as "An electronic application through which individuals can access, manage and share their health information, and that of others for whom they are authorized, in a private, secure, and confidential environment" [2]. The following benefits can be attained with PHRs [3][4]: (i) they provide a unified summary of users' entire health histories; (ii) they are easy to understand and use; (iii) 24/7 access to all users' healthcare data from anywhere in the world; (iv) collaborative disease tracking; and (v) continuous communication between patient and physicians.

At present, a number of cutting-edge companies such as Google and Microsoft attempt to deliver healthcare services with their own PHRs: Google Health and Microsoft HealthVault. However, a number of new security and privacy threats hang over patients' health data in this context [5]. Information might be fragmented and accessible from several sites (by visiting different doctors' offices, hospitals, providers, etc). Safety defects in some of these systems could cause the disclosure of information to unauthorized people or companies, and health data therefore need protection against manipulations, unauthorized access and abuses. Data needs careful

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 36–43, 2011.

protection, thus leading to the necessity for extreme strictness in storage and information exchange activities.

These threats are arguably more challenging than those found in most other industry sectors owing to [6]: (1) the amount of patient health record entries; (2) the number of healthcare personnel and organizations that might come into contact with a patient at any one time; (3) the difficulty involved in classifying the sensibility of a patient record entry; (4) the provision of very rapid appropriate access in a distributed computing environment; (5) the need for reviews of access permissions and for the PHR entries to be rigorously managed.

This research aims to study the privacy and security of PHRs from a usability perspective. The privacy policies of 22 free web-based PHRs were analyzed to extract their main characteristics. Our study on usable privacy and security is classified as a *security feature study*, according to the five-category framework presented by Birge [7]. The remainder of the paper is organized as follows. Section 2 justifies the importance of this research by reviewing the related literature. Section 3 introduces the method used in the experiment: instrumentation, experimental procedure and analysis of characteristics. Section 4 shows the principal results obtained from the data collected. Finally, Section 5 presents some concluding remarks.

2 Related Work

In recent years, relevant research has been carried out into PHRs, particularly with regard to the most popular PHRs, Google Health and Microsoft HealthVault. Martino and Ahuja [8] assessed these platforms to highlight those vulnerabilities existing in PHR privacy policy coverage and gaps in privacy policy notification mechanisms. The authors used well-researched evaluation criteria reported in literature [9, 10, 11].

Kotz et al. [12] compare existing privacy frameworks and identify a privacy framework for mobile healthcare and home-care systems. Mohan and Blough [13] propose a framework which supports the need to change the rule and policy combination algorithms dynamically based on contextual information. Williams [14] presents a survey of the research literature related to the security and privacy of personal health information in social networking applications, and provides a snapshot of privacy and security safeguards for social network websites. Huang et al. [15] design a method according to HIPAA guidelines to preserve the privacy and security of patients' portable medical records in portable storage media. Carrión et al. evaluate the privacy policies of 20 PHRs to check that the privacy of patients' data was preserved in accordance with HIPAA guidelines [16]. Sunyaev et al. develop criteria to assess the PHR systems from the users' viewpoint. The criteria were classified into three categories: patient information, personal control and additional services [17]. To illustrate their proposal, these authors applied the criteria to Google Health and Microsoft HealthVault. Sunyaev et al. [18] also performed an evaluation of Google Health API and Microsoft HealthVault API. An evaluation of the ethical, legal, and social issues (ELSI) was presented by Cushman et al. [19], who group this evaluation in four areas: privacy and confidentiality, data security, decision support, and the legal-regulatory regime for health data.

In this work, we analyze 22 free web-based PHRs, including Google Health and Microsoft HealthVault. To the best of our knowledge, no other studies have dealt with the security and privacy features of so many PHRs.

3 Classification Framework

3.1 Instrumentation

The problem of selecting the set of PHRs to be included in the study was confronted by consulting the web site of the American Health Information Management Association (AHIMA): *myPHR*. Its url is *www.myphr.com* and contains large amount of information on the PHRs. To the best of our knowledge, this web site provides the most comprehensive list of PHRs that a user can find, and has also been used to select PHRs in multi-source sampling [20]. It includes a section called “Choose a PHR” which was used to obtain a free web-based PHRs list. At the moment of accessing it on March 2011, a total of 29 free web-based PHRs (specified in Table 1) were retrieved. Note that AHIMA classifies the PHRs according to their format and cost. In this respect, some PHRs have premium accounts, and can thus also be classified as “for purchase".

Table 1. PHRs lists hosted in AHIMA

Format	Cost	Amount
Web-based	Free	29
Software-based	Free	0
Paper-based	Free	3
Web-based	For Purchase	63
Software-based	For Purchase	1
Paper-based	For Purchase	13

3.2 Experimental Procedure

The privacy policy of each PHR selected was reviewed by one author. This review was carried out between February and March 2011. Difficulties were encountered in seeking the PHRs' privacy policy because some of them were not on the PHR home page and others were fragmented in several documents. While the review was conducted, the formulation criteria for the enquiries was carefully discussed and agreed in an attempt to obtain a comparative framework that would be as comprehensive and clear as possible. Moreover, a number of the listed PHRs were not in force, or had no privacy policy or an equivalent document. In conclusion, the original amount of PHRs for participation (29) was reduced to 22.

The ISO/TS 13606-4 [6] is a Technical Specification (TS) which provides a basic framework that can be used as a minimum specification for an EHR (Electronic Health Records) access policy, and a generic representation for the communication of policy information. We have adopted ISO/TS 13606-4 as the basis of a classification framework for the evaluation of the set of relevant PHR features. According to the ISO/TS 13606-4: (P1) “health records should be created, processed and managed in ways that guarantee the confidentiality of their contents and legitimate control by patients in how they are used"; (P2) "the subject of care has the right to play a pivotal role in decisions on the content and distribution of his or her electronic health record, as well as rights to be informed of its contents"; (P3) "the communication of health record information to third parties should take place only with patient consent". A number of PHR security and privacy features have been defined according to these principles.

3.3 Analysis of Characteristics

Nine security characteristics were defined to analyze the PHRs: Privacy policy location, Data source, Data managed, Access management, Access audit, Data accessed without the user's permission, Security measures, Changes in privacy policy and Standards:

Privacy Policy Location. This characteristic is related to the question *Where is the Privacy Policy on the PHR web site?* PHRs should provide a Privacy Policy which describes how users' data are used in order for users to be informed. The Privacy Policy should be easily accessible by users. The difficulty of Privacy Policy access is assessed by counting the number of links clicked. The values that this characteristic may take are:

0. The Privacy Policy is not visible or not accessible.
1. The Privacy Policy is accessed by clicking one link.
2. The Privacy Policy is accessed by clicking two or more links.

Data Source. This characteristic is related to the question *Where do users' PHR data proceed from?* Generally, the user is his/her data source, but there are PHRs which do not only use this source. Some contact the users' healthcare providers, others allow other users and different programs to enter users' data and others use self-monitoring devices to obtain users' data. The values that this characteristic may take are:

0. Not indicated.
1. User.
2. User healthcare provider.
3. User and his/her healthcare providers.
4. User, other authorized users and other services/programs.
5. Self-monitoring devices connected with the user.

Data Managed. This characteristic is related to the question *Who do the data managed by the users belong to?* The users can manage their own data, but they can sometimes manage other users' data, such as that of their family. The values that this characteristic may take are:

0. Not indicated.
1. Data user.
2. Data user and his/her family data.

Access management. This characteristic is related to the question *Who can obtain access granted by the users?* The users decide who can access their PHR data. The PHR systems analyzed allow access to be given to different roles. The values that this characteristic may take are:

0. Not indicated.
1. Other users and services/programs.
2. Healthcare professionals.
3. Other users.
4. Other users, healthcare professionals and services/programs.

Access audit. This characteristic is related to the question *Can users see an audit of accesses to their PHRs?* The values that this characteristic may take are:

0. No.
1. Yes.

Data accessed without the user's permission. This characteristic is related to the question *What data are accessed without the user's explicit consent?* The PHR systems typically access certain data related to the users in order to verify that everything is correct. The values that this characteristic may take are:

0. Not indicated.
1. Information related to the accesses.
2. De-identified user information.
3. Information related to the accesses and de-identified user information.
4. Information related to the accesses and identified user information.

Security measures. This characteristic is related to the question *What security measures are used in PHR systems?* There are two types of security measures: physical measures and electronic measures. The physical security measures are related to the protection of the servers in which the data are stored. The electronic security measures are related to how stored and transmitted data are protected, for example, by using a Secure Sockets Layer (SSL) scheme. The values that this characteristic may take are:

0. Not indicated.
1. Physical security measures.
2. Electronic security measures.
3. Physical security measures and electronic security measures.

Changes in Privacy Policy. This characteristic is related to the question *Are changes in privacy policy notified to users?* Changes in Privacy Policy should be notified to users in order to make them aware of how their data are managed by the PHR system. The values that this characteristic may take are:

0. Not indicated.
1. Changes are notified to users.
2. Changes are announced on home page.
3. Changes are notified to users and changes are announced on home page.
4. Changes may not be notified.

Standards. This characteristic is related to the question *Are PHR systems based on privacy and security standards?* The PHR systems analyzed use or are based on two standards: the *Health Insurance Portability and Accountability Act* (HIPAA) [21] and the *Health On the Net Code of Conduct* (HONcode) [22]. The values that this characteristic may take are:

0. Not indicated.
1. HIPAA is mentioned.
2. System is covered by HONcode.
3. HIPAA is mentioned and system is covered by HONcode.

4 Results

Table 2 shows the results obtained for each PHR system included in the review. For example, Google Health takes the value 1 (according to the numbering in Section 3) for the PL characteristic. The scores were calculated according to the security and privacy characteristics of the PHRs. The following variables are considered in this study: The security level was quantified by employing the characteristics: Access audit and Security measures. The privacy level was quantified by using the characteristics: Privacy policy location, Cookies, Changes in privacy policy and Data accessed without the user's permission. Fig. 1 shows these scores as two overlapping histograms. In general, quite a good level can be observed in the characteristics analyzed. Nevertheless, some improvements could be made to current PHR privacy policies to enhance specific capabilities such as: the management of other users' data, the notification of changes in the privacy policy to users and the audit of accesses to users' PHRs, as shown in Table 2.

Table 2. Characteristics of each PHR system. PL: Privacy policy location; DS: Data source; DM Data managed; AM: Access management; AA; Access audit; DA: Data accessed without the user's permission; SM: Security measures; CP: Changes in privacy policy; S: Standards.

Tool	PL	DS	DM	AM	AA	DA	SM	CP	S
1. Google Health	1	4	1	1	1	3	3	2	1
2. ZebraHealth	2	1	0	0	0	1	3	4	1
3. myHealthFolders	1	1	2	2	1	1	3	1	0
4. Keas	1	4	1	0	0	2	3	3	0
5. EMRy Stick Personal Health Record	2	1	1	0	1	1	0	0	0
6. My HealthVet	2	1	1	2	0	1	2	0	1
7. myMediConnect	0	3	1	2	0	0	3	0	1
8. MyChart	1	2	1	0	1	4	0	0	1
9. MedicAlert	1	1	1	3	0	2	3	2	0
10. Microsoft HealthVault	1	4	1	4	1	1	3	2	3
11. MediCompass	1	5	1	2	0	2	3	0	3
12. TeleMedical	1	1	2	0	0	0	2	2	2
13. Health Butler	1	1	1	2	0	2	0	4	0
14. NoMoreClipBoard.com	1	3	2	2	1	2	2	2	1
15. MiVIA	1	0	1	2	0	3	3	2	1
16. iHealthRecord	1	0	0	0	0	1	2	4	0
17. Dr. I-Net	1	3	1	2	0	0	3	0	0
18. My Doclopedia PHR	1	2	1	2	0	3	2	2	1
19. dLife	1	0	0	0	0	4	2	2	0
20. RememberItNow!	1	4	1	4	1	3	2	3	0
21. MedsFile	1	1	1	0	1	4	1	1	0
22.Juniper Health	1	1	2	0	0	2	3	2	0

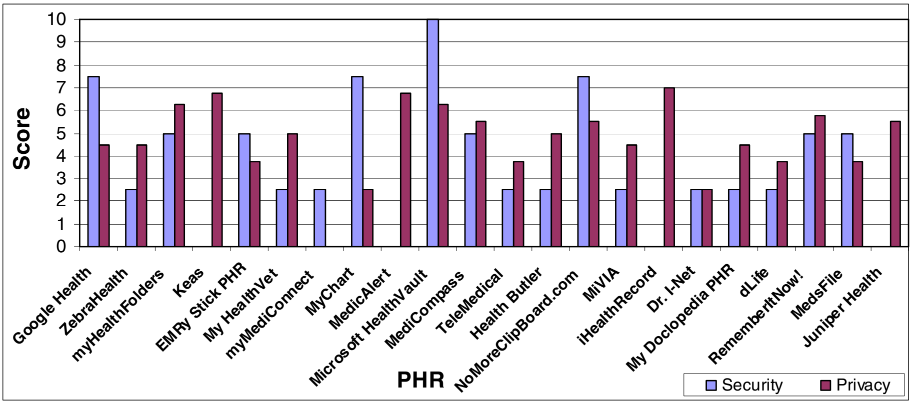


Fig. 1. PHRs usability score distribution

5 Conclusions

In this paper, a survey of 22 free web-based PHRs has been carried out. Most PHRs (86.36%) access different kinds of information without the users' explicit consent, but access to their identified information without their consent is never acceptable. Only thirteen PHRs (59.09%) notify or announce changes in their privacy policy to users. A number of PHRs (40.90%) do not provide any information about how the users provide access to their PHR data. In three PHRs (13.64%), physical security measures and electronic security measures are not reported. The Privacy Policies must include this aspect in order for users to perceive that their information will be protected by the PHR system. In a large number of PHRs (63.64%) the users cannot see an audit of accesses to their PHRs. This aspect should be improved because the users should be aware of how their information has been shared. A number of the PHRs reviewed (36.36%) are not regulated by a standard setting body. In future work, we intend to extend our analysis to a wider sample of PHRs, and compare the security and privacy features of free web-based PHRs with proprietary web-based PHRs.

Acknowledgments. This work has been partially financed by the Spanish Ministry of Science and Technology, project PANGEA, TIN2009-13718-C02-02.

References

1. Sood, S.P., Nwabueze, S.N., Mbarika, V.W.A., Prakash, N., Chatterjee, S., Ray, P., Mishra, S.: Electronic Medical Records: A Review Comparing the Challenges in Developed and Developing Countries. In: HICSS 2008. IEEE Computer Society, Washington, DC (2008)
2. Connecting for health personal health working group. connecting for health. the personal health working group final report (2003), http://www.providersedge.com/ehdocs/ehr_articles/The_Personal_Health_Working_Group_Final_Report.pdf
3. myMediConnect Personal Health Records, http://www.mymediconnect.net/phr.php

4. Tang, P.C., Ash, J.S., Bates, D.W., Overhage, J.M., Sands, D.Z.: Personal health records: Definitions, benefits, and strategies for overcoming barriers to adoption. J. Am. Med. Inform. Assoc. 13(2), 121–126 (2006)
5. Farzandipour, M., Sadoughi, F., Ahmadi, M., Karimi, I.: Security requirements and solutions in electronic health records: Lessons learned from a comparative study. Journal of Medical Systems 34(4), 629–642 (2010)
6. ISO/TS 13606-4. Health informatics Electronic health record communication part 4: Security (2010), http://www.iso.org/
7. Birge, C.: Enhancing research into usable privacy and security. In: Proc. of the 27th ACM International Conference on Design of Communication, SIGDOC 2009, pp. 221–226. ACM, New York (2009)
8. Martino, L., Ahuja, S.: Privacy policies of personal health records: an evaluation of their effectiveness in protecting patient information. In: Proc. of the 1st ACM International Health Informatics Symposium, IHI 2010, pp. 191–200. ACM, New York (2010)
9. Review of the Personal Health Record (PHR) service provider market: Privacy and security. ALTARUM Research (January 2007), http://www.hhs.gov/healthit/ahic/materials/01_07/ce/PrivacyReview.pdf
10. Detailed PHR privacy report cards. Patient Privacy Rights Foundation (2010), http://patientprivacyrights.org/detailed-phr-privacy-report-cards
11. User Centric (2010), http://www.usercentric.com/publications/2009/02/02/googlehealth-vs-microsoft-healthvault-consumers-compare-onlinepersonal-hea
12. Kotz, D., Avancha, S., Baxi, A.: A privacy framework for mobile health and home-care systems. In: Proc. of the First ACM Workshop on Security and Privacy in Medical and Home-care Systems, SPIMACS 2009, pp. 1–12. ACM, New York (2009)
13. Mohan, A., Blough, D.M.: An attribute-based authorization policy framework with dynamic conflict resolution. In: Proc. of IDTRUST 2010, pp. 37–50. ACM, New York (2010)
14. Williams, J.: Social networking applications in health care: threats to the privacy and security of health information. In: Proc. of the 2010 ICSE Workshop on Software Engineering in Health Care, SEHC 2010, pp. 39–49. ACM, New York (2010)
15. Huang, L.C., Chu, H.C., Lien, C.Y., Hsiao, C.H., Kao, T.: Privacy preservation and information security protection for patients' portable electronic health records. Comput. Biol. Med. 39, 743–750 (2009)
16. Carrión, I., Fernández Alemán, J.L., Toval, A.: Assessing HIPAA standard in practice: PHRs Privacy Policies., In: Proc. of IEEE EMBC 2011 (accepted for publication, 2011)
17. Sunyaev, A., Chornyi, D., Mauro, C., Krcmar, H.: Evaluation framework for personal health records: Microsoft healthvault vs. google health. In: Proc. of the 2010 43rd Hawaii International Conference on System Sciences, HICSS 2010, pp. 1–10. IEEE, Los Alamitos (2010)
18. Sunyaev, A., Kaletsch, A., Krcmar, H.: Comparative evaluation of Google Health API vs. Microsoft Healthvault API. In: Proc. of the Third International Conference on Health Informatics, pp. 195–201 (2010)
19. Cushman, R., Froomkin, A.M., Cava, A., Abril, P., Goodman, K.W.: Ethical, legal and social issues for personal health records and applications. J. of Biomedical Informatics 43, S51–S55 (2010)
20. Hulse, N.C., Wood, G.M., Haug, P.J., Williams, M.S.: Deriving consumer-facing disease concepts for family health histories using multi-source sampling. Journal of Biomedical Informatics 43(5), 716–724 (2010)
21. HIPAA (2010), http://www.cms.gov/HIPAAGenInfo
22. Boyer, C., Selby, M., Scherrer, J.R., Appel, R.D.: The health on the net code of conduct for medical and health websites. Computers in Biology and Medicine 28(5), 603–610 (1998)

Shining Chrome: Using Web Browser Personas to Enhance SSL Certificate Visualization

Max-Emanuel Maurer, Alexander De Luca, and Tobias Stockinger

University of Munich,
Media Informatics Group,
Amalienstr. 17,
80333 München, Germany
{max.maurer,alexander.de.luca}@ifi.lmu.de,
stockinger@cip.ifi.lmu.de

Abstract. Average users lack the technical expertise to understand SSL certificates and security is not their primary goal. Thus, it is very hard to create a notable impact on user behavior using SSL-status indicators. However, with the introduction of web browser Personas (simple skins) as a possibility to change the browser's chrome, it becomes possible to provide a large status indicator without wasting screen real estate. In this work, we present an evaluation of Personas to represent the current SSL status combined with newly designed SSL warning messages, both in the lab and in the field. Results suggest that the concepts positively influenced security awareness.

Keywords: SSL certificates, Security Awareness, Security.

1 Introduction

Communicating security issues to Internet users is a challenging task. One part of the problem space is how to visualize encryption state. A properly encrypted connection is an important security factor, even though it does not implicitly guarantee that the user is safe. In the past years the visualization in browsers has changed: The overlooked padlock icon has vanished and indicators in and around the URL bar have been introduced. The understanding of certificates has also become more complex with the introduction of Extended Validation certificates. With this kind of certificate, not only the match of encryption key and domain is verified but also certain company specific properties are validated [3]. This means that current indicators have to be adapted to display the difference between the two certificate types.

In this paper, we present SSLPersonas, which visualizes the current SSL status of the browser by changing the browser's chrome. Additionally, we propose a new type of warning pages that block website access for insecure SSL statuses. Both ideas were implemented as a plugin for Mozilla Firefox and evaluated in a lab study as well as using questionnaire results from 169 real users of the plugin (that had used the plugin for up to six months).

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 44–51, 2011.

Fig. 1. Proposed Personas for **a)** Extended Validation certificate **b)** Standard SSL certificate **c)** Warning (unencrypted content); **d)** Browser with a standard Persona displaying the changed warning page for a non-matching URL.

Visualizing certificate details has been recently covered by Biddle et al. [2]. They evaluated a new kind of interface for those details based on guidelines making it easier to understand the certificate related information and in that way improved the accuracy of security decisions in that way. Alternatively, the users can be actively interrupted during their browsing task by showing them a dialog box and forcing them to decide on an option [4]. Unfortunately users tend to get quickly habituated and dismiss those messages without paying further attention [1]

A big problem of web security is that it is not the users' primary goal [11]. Together with the phenomenon of change blindness [6] this results in passive indicators not being noticed.

2 SSLPersonas

Modern web browsers allow for completely changing the browser's chrome. Based on this, we came up with the idea of using this space for the visualization of the SSL certificate status. The expression “Persona” is used throughout this work as a reference to a special kind of browser skin, not changing the whole UI but only the window background. Personas therefore enable framing the current web page occluding a relatively big portion of the browser without wasting space for other UI elements in contrast to other security plugins or toolbars. By separating the browser chrome and the content, it is extremely hard for attackers to modify or mimic a Persona. The browser in figure 1d is using a standard Persona.

Most browsers already use colors in the URL bar to make people aware of the current SSL state. We combine this approach by using specially designed Personas that are displayed in the browser while visiting a website that uses an SSL certificate. The Personas use a similar color scheme as the Firefox browser itself and are additionally enhanced with some imagery to make their meaning more clear. A green Persona (see figure 1a) showing two padlock icons and a certificate icon is shown for Extended Validation certificates. For standard SSL certificates a blue Persona with only one padlock icon is shown (see figure 1b). For web pages being SSL secured but transmitting some information over unencrypted channels, a yellow warning Persona is used (see figure 1c).

Browsing SSL secured sites can additionally lead to blocking browser pages as recommended by the W3C [9]. In the case that a certificate has not been verified by a certificate authority – e.g. a self signed certificate – a warning message blocks access to this site. Another problem can occur when a certificate is used for a domain it was not issued for. We redesigned the upcoming warning messages by firstly modifying text and appearance and secondly by adding a preview image of the blocked site (see figure 1d) (without really loading the site).

3 Lab Study

3.1 Methodology and Participants

The concept was examined in a lab study by measuring if SSLPersonas is able to change the opinion of a user towards a specific website. We used a between subjects design with 24 participants. Twelve used a standard version of Mozilla Firefox, the other half had the plugin installed. The participants had to browse 14 different websites in seven categories. After each website, they had to report their personal assessment about security and trustworthiness of the website. We provided no explanations how the plugin worked or what the purpose of the study was until all assessments were made.

The 14 websites covered seven different cases. For each case, a well known and a hardly known web page was used – measured by their Alexa rating[1] for Germany, the country has been conducted in. The websites were shown to the participants in random order. The seven different cases consisted of the three different Persona states – Extended Validation (1), standard validation (2), mixed content (3) – and two kinds of non-SSL-secured sites – genuine (4) and Phishing ones (5) – and finally two different types of warnings – mismatched domain (6) and unknown issuer (7).

After viewing each website, participants had to answer a set of questions. For all Persona related websites, people were asked four questions: First they should tell whether they knew a website beforehand or not. This was used to check if our assignment for well-known and unknown websites did hold. After that, they had to rank the trustworthiness of the site on a five point Likert scale ranging from -2 (“this website seems suspicious”) to +2 (“this website seems trustworthy”). Another question concerned whether people would log in on the respective website -2 (“I definitely would not log in on this site”) to +2 (“I would log in without concerns”). The fourth question asked whether security assessment was possible: -2 (“I cannot see whether this site is secure”) to +2 (“There are enough indicators that this site is secure/insecure”). People should also name those indicators. Finally, we asked questions to determine how people would have reacted to the warning pages (cases 6 and 7).

The 24 participants – mostly students – were randomly assigned to one of the two groups. The 12 participants of the control group were in average 27 years old (ranging from 14 to 45; two thirds male). They used the Internet for 4.2 hours (SD 2.6) in average each day. All of them used the Internet for shopping and communicating; nine of them for online banking. Looking at the plugin group, demographics are

[1] Alexa.com computes a website ranking in a worldwide and country-based manner.

Table 1. Likert medians and means for the unknown websites of the different groups

	SSL						No SSL			
	Extended Validation (1)		Standard SSL (2)		Partially not encrypted (3)		Genuine (4)		Phishing (5)	
Trustworthiness										
Median	1,5	**1**	1	**0**	-1	**0**	0	**0**	-1	**-1**
Mean	*1,50	**0,83**	*0,67	**0,58**	*-0,50	**0,17**	*-0,33	**0,17**	*-0,58	**-0,5**
SD	0,52	**0,94**	0,89	**1,00**	1,24	**1,11**	0,89	**0,72**	1,56	**1,68**
Would login										
Median	1	**1**	-0,5	**-0,5**	-1	**0**	0	**-0,5**	-1,5	**-2**
Mean	*0,92	**0,33**	* -0,5	**-0,6**	* -0,5	**-0,2**	-0,4	***-0,5**	-0,8	***-0,8**
SD	1,16	**1,23**	1,38	**1,44**	1,38	**0,94**	1	**1,17**	1,4	**1,53**
Can determine security										
Median	1	**0**	-0,5	**-1**	0	**-1**	-1	**-0,5**	-1,5	**1**
Mean	*0,58	**-0,1**	* -0,1	**-0,7**	* 0	**-0,7**	* -0,9	**-0,7**	-0,8	***0,42**
SD	1,51	**1,08**	1,24	**1,44**	1,21	**1,23**	1,08	**1,3**	1,47	**1,38**

Secure | Insecure | * Plugin better | * Plugin worse | Plugin group | Control group

nearly equal: 58 percent male; average age of 23 years (range 14 to 30); using the Internet for 3.3 hours (SD 1.9) per day. Ten people use the Internet for shopping, twelve for communication and eleven for online banking.

Hypotheses: We formulated four hypotheses: **H1**: Using Personas for positive SSL statuses – like an SSL certificate being present – will increase the trustworthiness of a web page. **H2**: Using a Persona to warn users about a web page transmitting some data over unencrypted channels will reduce the trustworthiness. **H3**: The absence of the Persona on non-encrypted websites will reduce their trustworthiness. **H4**: The redesigned SSL warning pages will enable more users to choose the correct/secure option.

3.2 Results

The classification of known and unknown sites worked well for cases 1, 2 and 5. The websites for case 3 and 4 were known by hardly any participant. In contrast, the classification of unknown sites matched very well. Thus, the following results will mostly be taken from the unknown websites. For the well-known websites, we found a tendency of them being rated nearly equally to the unknown sites whilst the plugin seemed to have more influence on people using unknown sites.

Comparing the five cases, the SSLPersona-enhanced browser outperformed the standard one. The medians and means of the participants' answers ranging from -2 to +2 can be found in table 1. In case of correctly SSL-secured sites (case 1 and 2), people in the plugin group rated trustworthiness, the willingness to log in and the ability to determine security higher. These values support H1.

For the partially encrypted page using the warning Persona (case 3), we expected to get lower ratings for trust and log in willingness. The ability to determine security should still be higher. Again all those assumptions hold when looking at the median values. This supports H2.

In case of the non-SSL sites and the phishing sites, both groups saw a standard any Persona. For H3, we expected the results of the plugin group to drop due to a missing

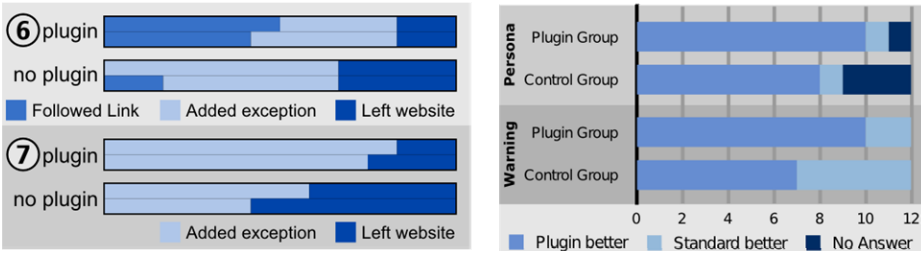


Fig. 2. Left: Warning message decisions for cases 6 and 7. Upper half of each bar represents well-known websites, the lower half hardly known websites. **Right:** When showing participants pictures of the standard or plugin enhanced version, people from both groups preferred the enhanced version.

positive feedback. This did not happen. The study duration was much too short for people to get used to the plugin and expecting the green Persona to show up. Furthermore, the order of the visited websites was randomized for each participant.

Looking at the warning page for a non-matching URL (case 6), participants had three options to choose from: Leaving the site, visiting the link the certificate was intended for and adding an exception and continuing to the site. The best option in such a case would be to visit the URL the certificate was originally issued for. This was done by zero (known) and two (unknown) participants for the control group in contrast to six (known) and five (unknown) using the plugin (see figure 2 left).

When confronted with an untrusted issuer certificate (case 7), the only option was to leave the site or set up an exception. Since our websites both were genuine but used self signed certificates, setting up an exception was okay in this case. Again plugin users used the exception more often (see figure 2). Comparing the correct answers with a repeated-measures ANOVA – within-subject factors: “known” and “warning type” – the results for the between subjects variable “group” are highly significant ($F_{1,22}=16$, $p=0.001$) which confirms H4.

After rating the websites, the participants were debriefed and shown some side-by-side images of a browser with and without the plugin. Overall, 75% of the participants preferred the plugin version. The same holds for the comparison of the warnings. 71% preferred the plugin warnings (see figure 2 right).

4 Field Study

Based on the lab study, the plugin was improved (e.g. UI bugs fixed) and published on August 20, 2010 through the Firefox addon webpage. It quickly gathered public interest and was installed several thousand times. Different blogs and podcasts related to security covered the plugin and reported about the concept [5,8]. By March 2011, the plugin was downloaded more than 15,000 times. It has around 2,300 active users. 86% use Windows, 8% Unix and 6% Mac OS. 45% percent are US English users, 32% are German, 5% British followed by smaller amounts of French, Spanish, Italian, Brazilian and Russian users.

4.1 Methodology and Participants

Having a real-world user base with a broad spectrum of nationalities provides a good baseline for evaluating experiences with the plugin and its concept. Thus, we created an online questionnaire and a plugin update to invite them to take part in the research. No further incentives were provided.

The questionnaire consisted of three sets of questions. Firstly, a small set of nine questions from the IBM “Post-Study System Usability Questionnaire” [7] was used with 5-point instead of 7-point Likert scales. The second set contained eleven questions about plugin usage and security knowledge. Finally, we collected demographics of the participants.

The survey was available in English and German. Incoming participants were diverted according to their browser language. The survey was also available through several external links – like the plugin's web page.

After deploying an update, Firefox automatically recommends it to the users. Therefore, most of the users quickly updated (approx. 1,700) in the two weeks period the questionnaire was available. From those users, 169 (approx. 10%) completed the questionnaire. 88% of them used the English version 12% the German version. The average age of the participants was 41 years (range: 9 to 70). Only 16 participants (9.5%) were female. Please note that there was no way to check whether the demographic data was valid.

4.2 Results

Analyzing the responses, we found that a large number of respondents has to be categorized as sophisticated users. In average, participants spend 30 hours (SD 23.1) per week on the Internet, 12% more than 60 hours. Rating their computer skills, only 2% stated to be “inexperienced”, 21% had “average” skills, 48% had “advanced” and 28% “expert” skills. 9% of the participants stated they had been victim to a Phishing attack before, but only one attack successfully fooled the victim.

We had no technical means to find out how long people had installed the plugin before and added a respective question. 15% stated having used the plugin less than one month, 40% 1-2 months, 30% 3-5 months and 15% more than five months. Comparing the groups of new and old users, no major changes in answers can be found except for the stated experience level. Whilst 69% of the long-term users stated either to be advanced or expert Internet users, those users made up only 48% of the group of new users. Probably early adopters were rather experts whilst more and more inexperienced users discovered the plugin lately. The high number of expert users can be explained with the effort of finding an extra security plugin which is usually not taken by novice users due to the secondary goal property of security [10].

In the questionnaire, we showed pictures of Firefox's standard SSL indicator – a green bar in front of the URL indicating an Extended Validation certification. We asked them whether they had noticed it before. 82% confirmed this. We asked those users to explain what it meant: 11% gave a correct explanation (mentioning the concept of Extended Validation) and 57% mentioned at least security. From the 82% of people that stated they had noticed the bar, 50% said they had already clicked on this field at least once and half of those were able to correctly describe the contents of

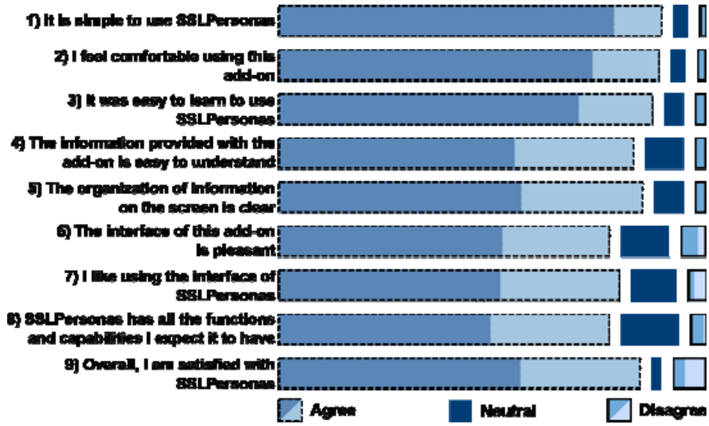


Fig. 3. The distribution of Likert answers throughout the nine questions extracted from the IBM Post-Study System Usability Questionnaire

this field. “Do you know the difference between the blue and green Persona.” resulted in 62% saying they knew it but again only 21% gave a correct answer. Another 21% of the respondents thought it had to do something with the “strength of encryption” and some people thought the blue Persona would indicate that some contents of the website are not encrypted. This means that although users estimated themselves to be security experts the detailed answers were mostly incorrect.

The IBM usability questions used a 5-point Likert scale ranging from 1-‘strongly agree’ to 5-‘strongly disagree’. All questions were answered with a Median of 1, which means that the plugin is very helpful to the users. The best mean value was for “It is simple to use SSLPersonas” having 1.2 (SD 0.6). The worst mean value was 1.7 (SD 0.9) for “SSLPersonas has all the functions and capabilities I expect it to have”. More details on the nine questions can be found in figure 3.

To our surprise, 38% stated that SSLPersonas changed the way they use the Internet.

5 Limitations, Discussion and Conclusion

As with any lab study, this one also had problems with ecological validity. The field study was supposed to fill this hole. Interestingly, we found that the real users of SSLPersonas largely consist of expert users. Although the concept primarily targets novice users, expert users dominate both, the early adopter and the basic user group while novice users adopt the concept much more slowly. We assume this is due to different factors. For instance, novice users do not know about the security issues and thus do not search for solutions. Therefore, a novice user would not install the plugin by herself. This suggests that especially security enhancements should be delivered with the browser.

Despite the high number of experts, the percentage of people that were able to correctly distinguish between the different kinds of SSL certificates – the green and the blue Persona – is very low. The number of people misleadingly assuming a stronger encryption is equally high. This puts the whole concept of different types of certificates into question as they seem not to be self-explaining. Our green Persona potentially added up to this fact because it contained two lock icons instead of one in the blue Persona. After the study, we changed that and added an icon of a "green man" as a metaphor for the validated company identity.

Both studies support our hypothesis that SSLPersonas improves security awareness in standard web browsing tasks without occupying extra screen space. In the lab study, we could show that SSLPersonas influences the users' rating of websites towards a better (more secure) behavior. The published plugin was installed by many people of which 169 participated in a survey. The results attest SSLPersonas positive effects on security-aware behavior. Being able to positively influence the users security decisions at no cost in combination with the fact that expert users also rely on the concept shows the value and importance of this work.

We also proposed new layouts for certificate warning pages with the main goal to avoid habituation. They enabled the participants to make significantly more rational decisions when dismissing the warning dialogs.

Personas as a status display for SSL information seem to be able to influence security assessment. In the future other factors could be added to the underlying concept to get a more complete security analysis.

References

1. Amer, T.S., Maris, J.B.: Signal words and signal icons in application control and information technology. Journal of Information Systems 21 (2006)
2. Biddle, R., van Oorschot, P.C., Patrick, A.S., Sobey, J., Whalen, T.: Browser interfaces and extended validation SSL certificates: An empirical study. In: CCSW 2009. ACM, New York (2009)
3. CA/Browser Forum: Extended validation ssl certificates, `http://cabforum.org`
4. Egelman, S., Cranor, L.F., Hong, J.: You've been warned: an empirical study of the effectiveness of web browser phishing warnings. In: Proc. CHI 2008, pp. 1065–1074. ACM, Florence (2008)
5. Gibson, S., Laporte, L.: Security now episode 277, `http://www.grc.com/securitynow.htm` (visited March 27, 2011)
6. Grimes, J.: On the failure to detect changes in scenes across saccades. Perception 2, 89–110 (1996)
7. Lewis, J.R.: IBM computer usability satisfaction questionnaires: psychometric evaluation and instructions for use. International Journal of Human Computer Interaction 7(1), 57–78 (1995)
8. Morton, B.: SSLPersonas - SSL blog - Entrust insights, `http://ssl.entrust.net/blog/?p=321` (visited March 27, 2011)
9. Roessler, T., Saldhana, A.: Web security context: User interface guidelines (2009)
10. Whitten, A., Tygar, J.D.: Why johnny can't encrypt: A usability evaluation of PGP 5.0. In: Proc. USENIX 1999, pp. 169–184 (1999)
11. Wu, M., Miller, R.C., Garfinkel, S.L.: Do security toolbars actually prevent phishing attacks? In: Proc. CHI 2006, pp. 601–610. ACM, New York (2006)

Ambient Displays: Influencing Movement Patterns

Tasos Varoudis

`arch+ech` architecture + technology,
Lab 24, Cremer Business Centre, E2 8HD, London, UK
varoudis@archtech.gr

Abstract. Ambient displays are gradually augmenting the principal static elements of architecture, such as walls, transforming space into a dynamic and ever-changing environment. Does the addition of such digital elements influence people's perception and understanding of space around them? If so, do ambient displays lead to behavioral changes like people's movement in such environments? In this particular study, a series of experiments were conducted to investigate public interior spaces with embedded ambient displays. The findings are then presented showing how the presence of an ambient display through its visual depth affects and changes movement patterns. This study discusses the ability of an ambient display to refine navigation paths and suggests that its visual depth can enhance its effectiveness.

Keywords: Ambient displays, human navigation, built environment, visual perception.

1 Introduction

Digital technologies have now become a medium for transforming the principle elements of architecture into dynamic components that affect people's perception and understanding of the space around them. Ambient displays, in most cases, can simply augment walls and thus have an advantage of not changing the physical morphology of space. Through light and sound, they make use of the entire physical environment as an interface and source to digital information [1]. This interface can be considered as a virtual opening, in a form of a digital window that extends the architectural space and widens the vision of users in a given space. Like a window that connects two physical spaces and transfers information between them, an ambient display can be seen as a link between physical and virtual worlds, where the virtual world can be for example the projection of another physical world or a computer generated space.

The introduction of a virtual space into the physical environment redefines the spatial properties of this environment. For instance, an ambient display can be used to transform a static wall to a more transparent and fluid element that dissolves the boundaries between virtual and physical space. During 'out of bounds' interactive installation, Chris O' Shea [2] effectively challenged the relation between physical and virtual space. By projecting previously captured spaces onto the wall, he enabled people to 'see' through the physical boundaries of space. As he stated, this interaction allowed people to enter the 'prohibited' areas of the museum while encouraging their

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 52–65, 2011.

childlike curiosity [2]. But how can ambient displays affect people's movement in space? The ability of ambient displays to project views of one space from another and create hybrid information space might alter our perception of space. If so, can the use of hybrid space create behavioral changes related to people's navigation in space?
The aim of this research was to analyze how digital information when introduced through ambient displays affects people's movement in the interior of public architectural spaces. The research's interest is in how the addition of an ambient display at a given space can alter people's navigation in it. In order to answer this, a series of experiments were designed to explore the effects of ambient displays on people's movement inside public spaces.

2 Background

Research from two fields is relevant to the work. First, this research examines work investigating the use of ambient technology in architectural spaces. Second, there is a brief outline of research studying perception and movement in architectural spaces, where ambient displays are not considered as elements that augment vision.

2.1 Ambient Displays: Transforming Architectural Spaces

The paradigm of ubiquitous computing represents the technological tendency of embedding information processing into everyday objects and activities. Drawing on Weiser's proposal of 'ubiquitous computing' [3], Ishii and Ulmer [4] envisioned the transformation of architectural surfaces into active interfaces between physical and virtual worlds. One of the first attempts to place an ambient display inside architectural space was initiated by Ishii et al. [5]. They designed the ambientROOM, a room equipped with ambient media (ambient light, shadow, sound, airflow, water flow). AmbientROOM provided the inhabitance of a room with a subtle awareness of human activity around people. The environment created by the ambientROOM was designed to enhance the sense of community through shared awareness concluding in a more pleasant space.

Currently, ambient displays have received considerable attention by architects who try to construct buildings with embedded digital technologies. For example, the Porsche Museum by Delugan Meissl [6] (see Figure 1) and the BMW Museum in Munich, Germany that was designed in 2007 by ART+COM [7] include large ambient displays that have augmented the interior surfaces of walls. In addition to this, when they designed Digital Pavilion in 2006, Kas Oosterhuis and Ilona Lénárd [8] transformed concrete interior walls into interfaces able to display readable information or create atmospheric lighting effects. Taking the idea of placing ambient displays instead of walls further, Ataman et al. [9] proposed the use of large display 'materials' as construction surfaces in architectural design. They envisioned movement through space in the future to be more dynamic, incorporating different levels of transparency and space that could be described by the fluidity of the walls that surround people.

Fig. 1. Porsche Museum, Stuttgart, Germany

As indicated by previous studies [10, 11], the presence of an ambient display in architectural space can ensure more pleasant environments; while at the same time can be informative and socially engaging. Moreover, because of being blended with physical space, ambient displays are unobtrusive and do not distract the users that are not interested in the displayed content. For example, Mathew and Taylor [10] with their project AuralScapes try, with the use of sound and sky projections, to create a virtual connection from the outside to the inside, where information is abstracted and delivered at the periphery of human attention. By developing this link, they create a pleasant indoor environment and partially dissolved the static notion of the surrounding walls.

In fact, the use of digital technologies in architecture goes beyond the simple modification and transformation of space into a pleasant and informative environment, influencing people's behavior in it. With his work Röcker et al. [12] states that ambient display installations that promote awareness and presence produce positive effects and behavioral changes on office teams.

It has been observed that changes to the digital environment have led to behavioral changes such as the movement of people through that space. Indeed, Schieck et al. [13] in an attempt to analyze the influence of an interactive floor installation in people's social engagement in urban environments, reports that she recorded unexpected and diverse changes in movement patterns around the installation. But what are the changes in movement patterns resulting from the presence of ambient technology?

For architects wanting to incorporate ambient interfaces it is essential to acknowledge the effects of their design proposals on people's navigation in space. With analysis and visualization of these effects, ambient display components will be incorporated in designs efficiently and activate or transform existing spaces. Relevant studies [6, 14, 15, 16] on ambient technology place the center of attention on human-computer interaction without considering the visual and spatial perception that link people and technology.

2.2 Perception and Movement in Architectural Spaces

It is generally accepted in architecture that the structure and configuration of space affect people's navigation and movement. Gibson's research, which was primarily developed for visual perception, suggests that our senses provide us with direct awareness of the external world and its properties [17]. People perceive space through their senses and act accordingly, thus there is a tight relation between perception and movement.

Based on this theory, architect and virtual reality pioneer Benedikt [18] proposed that space is perceived as a collection of visible surfaces that are not obstructed by physical boundaries and he defined 'isovists' to describe the area in the environment that is directly visible from a location within space. A single isovist is the area of space directly visible from a given location in space, together with the location of that point. For example, in a convex space or a rectangular space with partitions the isovist area of a given point may not include the full area of that space and some parts of the space will not be directly visible from other points in space (see Figure 2).

Another urban and architectural theory that is relevant to this study is 'Space Syntax'. Hillier and Hanson proposed 'Space Syntax' to describe and analyze the character of a space and its effects on human behavior [19]. 'Space Syntax' research shows that the majority of human movement occurs along the longest lines of sight, and that the more open visible space we have in front of us the more we tend to move towards that direction [20]. However, the complexity of the spatial elements that are taken into consideration is limited. 'Space Syntax' sees space as a set of solid walls and empty openings and does not examine transparent elements. In addition to the lack of consideration of transparent materials, there is also a lack in understanding the effects of ambient technologies and 'digital' transparencies. Both physical and 'digital' transparencies may have important effects on people's perception of space and movement within it as they both extend and sometimes distort the depth of field.

Existing spatial analysis theories do not take into consideration complex architectural components such as ambient displays. Despite the fact that there are

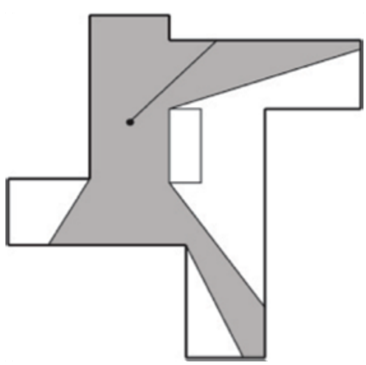


Fig. 2. Isovist field: grey color indicates the visible area from point

some studies that deal with pervasive systems in urban environments, they focus on social behavior and do not consider the influence of ambient displays on human movement [13, 21]. Additionally, such research is limited in considering ambient technology as a layer that is placed over existing urban infrastructures or simply replaces building facades. As we move into a world where a fusion of virtual and physical is going to be prevalent [22], studying and analyzing people's behavior in relation to the use of embedded ambient displays can offer important knowledge.

3 Aims and Objectives

The study starts with the assumption that the topological and visual relations between physical spaces are two important factors that determine the distribution of people's movement in space [23].

This research intends to analyze the changes in movement patterns when ambient displays are used as a virtual extension of the visual boundaries inside public architectural spaces. The hypothesis of this study is that placing ambient displays that virtually link and extend one physical space towards another that is near but not directly accessible will influence the topological and visual relations between spaces and as a result will affect the distribution of people's movement (see Figure 3).

Analyzing the effects of ambient displays on movement patterns when are blended in architectural spaces and induce virtual augmentation would provide a better understanding of the relation among humans, space and ambient displays. In contrast to current stable physical spaces, the introduction of new responsive and ever-changing materials will impose new fully adaptive architecture. Therefore, extending current knowledge and theories to involve digital transparencies as a crucial element of spatial configurations is inevitable.

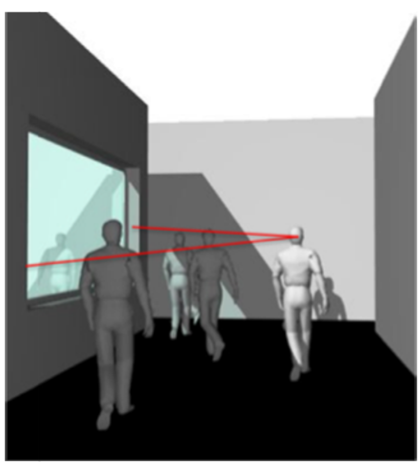


Fig. 3. Ambient display as digital transparency

4 Methodology

To achieve the aims of this research, a series of experiments were developed in order to observe and analyze people's movement in different spaces depending on the presence of ambient displays (see Figure 4). The experiments are focused on the use of ambient displays in the interior of public architectural spaces by placing them near a place of common interest. The main reason behind the use of public architectural spaces is that public spaces offer great opportunities for experimenting with ambient technology and the analysis of movement. From a methodological perspective, such places allow observing the movement behavior of a significant amount of people providing a larger sample for experimentation and analysis.

A fundamental form of space, in which it is simple to examine the flow and direction of people between discrete routes, is corridor-like settings. Based in this setting, two distinguished routes are needed, from which the users can choose in order to access the target space. 'Target space' is considered as a space with common interest for the participants of the experiment such as a coffee area or a common room that they can prepare and eat their lunch.

This type of experiment gives the opportunity to examine a single 'decision making' point providing clear and unbiased experimental conditions. It is essential that the routes need to be symmetrical in order not to influence the users' choice by producing different visual triggers. Moreover, having the 'T' shaped corridors as the setting, where people can go left or right, gives symmetrical visual characteristics. Therefore, the addition of an ambient display just before the corner as depicted in Figure 5 is expected to act as a visual trigger. The display's position near the corner will potentially introduce a digital opening in the visual field and will influence people's movement.

Fig. 4. Example of an experiment setting. Ambient display embedded in wall to the left.

Fig. 5. Influencing direction of movement. Ambient display embedded in wall to the right.

Intuition and a simplistic interpretation of ‘Space Syntax’ suggest that the presence of an ambient display in a corridor should not influence the route decision-making choices of occupants going down this corridor. In the condition when both routes are equidistant from the objective, one might well expect a 50/50 left-right split of occupants. Alternatively, the hybrid space hypothesis suggests that a digital ambient display that functions like a virtual window will alter spatial morphology and so will result in an observable behavioral change in movement patterns (see Figure 5). To test this, a set of corridors as a ‘T-shaped’ arrangement was constructed with the upper part of the ‘T’ near the target space. As the ‘target space’, a room for coffee providence and preparation was used, since it is more attractive for daily visits for both staff and visitors. The research’s interest is to examine how the addition of an ambient display would affect this distribution, a case that currently is not predictable by relevant theories like ‘Space Syntax’.

Having set an outline of the experimental requirements, the design phase divided in three parts. The first part is focused on the ambient display and its content while the second part on spatial setting. The third part is a combined effort for the effective placement of the ambient display in the particular space.

Deriving from the background research and the hypothesis, the display’s main function was to act as a live video link providing a one-way video from the ‘target space’ and introducing an augmented opening that digitally expands vision. In addition, a series of ‘null-tests’ were introduced as part of the experiment using the ambient display with: no content, random static images and random animated content. The ‘null-tests’ were critical for establishing the content of the display as the only source of influence.

4.1 Experimental Set-Up

The space used was the Ambient Technology Lab at the Open University UK. It is a space of approximately 144 m2 that is free of internal walls and can be configured in any arrangement easily with little limitations. For the construction of the corridor-like setting lightweight solid partitions were used covering the full height of the space and positioned in order to produce a symmetrical space. The remaining space at the end of the corridor-like setting was designed as a small cafeteria with sufficient space for people to stay and chat with each other reproducing a casual everyday atmosphere. Over the course of the experiment there was free coffee, tea and some sweets in this space available to everyone.

For the data gathering, a multi-camera set-up was used for synchronized recording of every corner of the lab. The cameras were positioned in the ceiling and configured to track multiple angles of the decision-point area as well as the full length of the corridor and the coffee area. Additionally, the cameras could also focus on the face along with the body of each subject in order to recognize the line of sight, as each subject was moving towards and passing the decision point. For every day of the experiment eight video streams were recorded during the working hours.

The experiment ran over a period of several weeks where people that work in the university or visitors could use this setting for taking their morning or afternoon coffee. None of the participants were aware of the research's nature but were informed that they were taking part in an experiment and would be recorded. The experiment area was clearly marked to the effect that they are entering a video monitored zone and emails sent out to ask for participants. The emails explained that the purpose of the experiment will be revealed after its compilations and the whole process was ethically approved. While participants were asked to use a specific entrance and exit door in order to avoid passing through the corridor in both directions, it was felt that this did not bias participants' response of direction choice.

As discussed before, the measurements had to be compared with a 'base model' of this space with no ambient displays installed. On that account, the experiment started by collecting data about people's movement within the T-shape configuration with no ambient display present. After having sufficient data to serve as the basis of the experiment, the second phase started, in which a large display was carefully embedded into the right wall (see Figure 6) of the constructed corridor just before the corner (decision point). The display, a large anti-glare display with a 15 mm black casing, was linked with a high-definition professional camera and depicted the coffee area as seen from a particular 'perspective'. This 'perspective' view was used in order to emulate the actual perspective of the coffee area that an opening at this position would have revealed. Similarly, in the third phase the display was placed symmetrically on the left side of the corridor, positioned in the same way and producing the same effect but streaming data of the coffee area as seen from the left side. Everything else in the experimental setting remained unchanged for the course of the experiment. At the end of the study all video streams were examined regarding the movement, choice of direction and the reactions of the people. All data were treated with confidentiality and not shared with anyone outside the limits of this research.

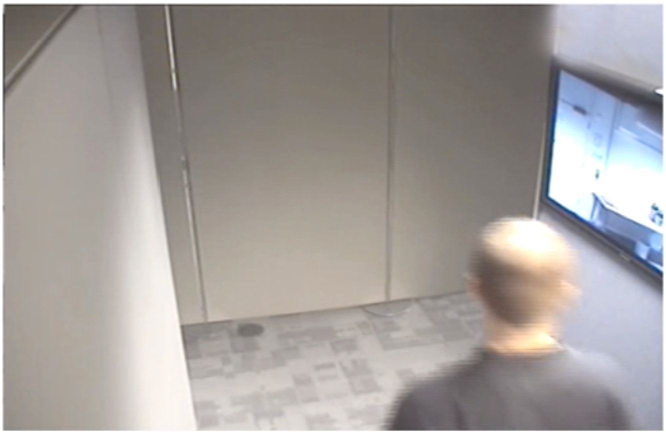


Fig. 6. Example of an experiment setting. Ambient display embedded in wall to the right.

4.2 Initial Reactions – Video Observation

To capture people's initial reactions to the ambient display the recorded video, which was taken from cameras tracking people's movement while approaching the decision point, was observed. Upon approaching the display a small percentage of people were seen to momentarily stop and look at it for just a second and then turn towards the route of the screen. The majority of people only seemed to take a glimpse of the screen by turning their head slightly and then turn at the corner. Although the video revealed that most of the individuals took their decision quickly, some individuals changed their initial decision and finally moved towards the side of the display. Overall, while the ambient display was something new to their environment less than 0.01% of people who took part in the experiment seemed unfamiliar to its presence. In total, more than 800 individuals took part in the experiment, which gives us a clear evaluation of the hypothesis. The number of participants doesn't include people passing multiple times but the observations are kept in order to later check their potential contribution on the overall hypothesis.

5 Results

To analyze the movement patterns in the experimental setting the number of individuals and groups walking through the corridor were counted in relation to their decision to follow the right or left direction in the particular setting. The data were categorized according to the experimental phase (without display, with the display on the left and with the display on the right) and whether it was individuals or groups. The analysis conducted by grouping the data into two categories in order to eliminate any signs of internal influence within the groups of people ('groups' have more than one person and all subjects follow the first person in the group): 1) for 'all groups and individuals' without taking into account if a person was alone or a part of the group and 2) 'only individuals' counting all 'individuals' and each of the 'groups' as one subject unaffected by the number of people in the group. These categories were analyzed using a chi-squared test and logistic regression analysis (modeling of binomial proportions).

In the first phase (no display), which took place over a period of four weeks, the findings revealed that combining groups and individuals, 55% of the people turned left in the specific setting and 45% turned right, while counting only the individuals the distribution was 54.8% left and 45.2% right (see Figure 7).

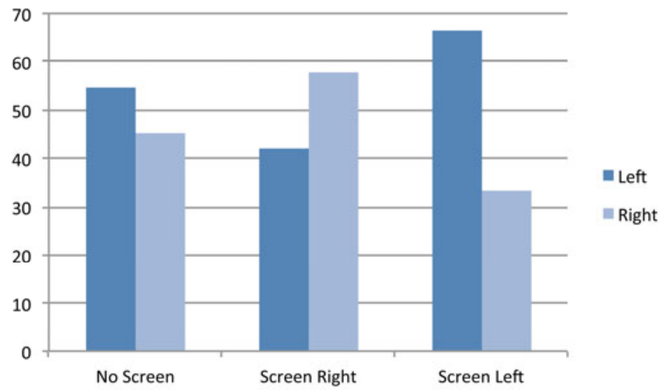


Fig. 7. Distribution (%) between routes

In the second phase where the display was on the right side and run over a period of three weeks, the distribution was 41.1% turning left and 58.9% turning right for the combined test subjects and 42.1% left and 57.9% right for only the individuals (see Figure 7). Those results showed a significant 'shift' in people's distribution along the two alternative directions. The shift towards the right side was 13.9% combining all test subjects together and 12.7% counting only the individuals.

Finally, in the third phase with the display positioned on the left side and run for three weeks, the distribution was 73.4% turning left and 26.6% turning right for the combined test subjects and 66.6% left and 33.4% right for only the individuals. Those results also showed a significant 'shift' in people's distribution (see Figure 7). The shift towards the left side was 18.4% combining all test subjects together and 11.8% when counting only the individuals.

Furthermore, for validation of the significance of the results from the three unmatched groups a chi-square statistical test was used. The p-value of this test was 0.000047. Figure 8 depicts the calculated 'expected' percentages from the chi-square test against the observed percentages.

Following the tests about the general significance of the measured data and the influence of the ambient display, a logistic regression analysis was conducted ('modeling of binomial proportions'). For this test a value denoting the position of the ambient display was introduced (direction of influence): '-1' when the ambient display is at the opposite side than the one testing, '0' when there is no display and '1' when the display is on the same side, as well as a value denoting the presence of an ambient display regardless of position (encoded as '0' for not present and '1' for present in one of the walls). The 'direction of influence' matches the direction of the significant change in movement as measured in the experiment, with a p-value (t pr.) less than 0.001. The result of this analysis further confirms the hypothesis that the ambient display produces a significant change in the direction of movement towards the side at which the display is located. Additionally, combining the 'direction of influence' and the 'presence' variables evinces that the resulted distortion is not based on the side of the corridor, where the ambient display is located, and thus for both 'right' and 'left' cases the change was equally significant.

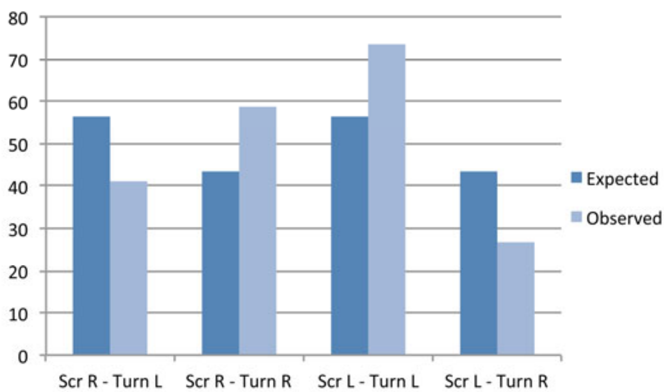


Fig. 8. Expected and observed (%) movement distribution

5.1 Null-Tests Results

The results presented until now showed that the presence of an ambient display influences how people move through the environment. However there was a need for extracting whether the ambient display itself (the device) or its content were responsible for that change. For that reason, using the same equipment and spatial setting a follow-up experiment was conducted positioning the display in the same places as the first experiments but without the video link. Instead of the video feed, several scenarios were tested: no image is used in order to test the ambient display as a device while static images, animated images and short video clips used to explore the effects of diverse content that could affect the users.

The results were treated in the same way as the first experiment and showed that the shift in direction between the phase without a display (base model) and each of the null-test phases is less than 3% in all cases and not statistically significant.

In detail, from more than 400 individuals who took part in the null-test experiments, the majority of people continue to use the left route more (the same way as before the application of the ambient display), with the percentage varying from 57% to 53% in favor of the left route. More analytically, compared to the phase without an ambient display where the distribution was 55% left and 45% right, the distribution with the display on the left for the worst case null-test scenario (case with the largest shift) was, 57.8% left and 42.2% right; while with the display on the right, 53.2% left and 46.8% right. What is clearly visible from the results is that the ambient display blends in the environment and becomes an unobtrusive object that without the visual extension does not influence people's behavior and movement in space.

6 Conclusion

The results of this study reveal that, an ambient display that shows typical information has, as one might expect, no change on pedestrian route choice behavior. If the display shows a projection of nearby space, relevant to the pedestrian, then the presence of the augmented visual depth and more particularly the position of the ambient display influences route choice behavior. This has a number of ramifications in terms of both the design of ambient displays in architectural settings and the use of augmented/hybrid spaces in research conditions.

Combining the findings from the experiment and the observation from the recorded video it is concluded that there is a significant change in the distribution of people and their movement patterns when an existing architectural space is augmented by the presence of an ambient display projecting a virtual link between two physically disconnected spaces (where the one can be also a virtual representation of a space). It is also clear from the results that the position of the display in the symmetrical routes has no different effects regarding the distortion of movement. In addition to this, the results of the 'null-tests' addressed that when the ambient display is not producing an augmentation of the visual field or the visual depth, it does not influence people's movement in space.

This study also demonstrates that the position of the ambient display as a visual link to a near but physically disconnected location has the effect of increasing awareness about this space giving a subliminal direction towards the side where the screen is situated. In some cases the ambient display unconsciously nudged people to pass through the corridor setting and go to the coffee area. As the video revealed, most of the users took decisions quickly. However, there were cases of individuals changing their initial decision and adapting their route towards the side of the screen. It is speculated that raising awareness of this decision may have increased the likelihood that these individuals recalculated the new augmented layout of space against the old and moved accordingly.

Accepting that people's visual perception of space becomes influenced by the virtual augmentation and extension of space that ambient displays produce, this research can contribute in two main areas. One area deals with the architectural space design featuring ambient displays, as well as navigation in space. The second area is the content and interface design for ambient media. In the fast-growing field of digital augmentation in architecture, understanding and acknowledging people's movement, proximity and navigation in space can give new ways of managing and directing movement towards desired places or interfaces. Examples within this area include subliminal nudges for accessibility of remote or 'hidden' spaces as well as alternative and more efficient methods to assist way-finding. In addition, ambient information systems can be made and positioned in a way able to enhance interactivity and social engagement.

In the field of human-computer interaction the design of the interface is of crucial meaning for the effectiveness of the ambient displays. As it is clear that people react to the virtual depth of the ambient interface, as indicated by the change in movement within the experiments, designing such interfaces needs a significant consideration of the results that they will induce. As this study shows, an ambient display that extends the visual depth acts as a subliminal element attracting people to approach it and also change their direction. This effect is clear when comparing the findings from the 'null-test' with the main experiment.

Further research is needed, however, to produce more generic insights that might help with ambient display usage in architectural space and more critically the interface design itself. For this reason, it is crucial to explore more areas of the visual augmentation of spatial boundaries with ambient displays. Currently, a research that explores how the directionality of the virtual perspective projection, which is produced by ambient displays, can influence people's direction of movement and proximity from the display is in its final stages. In essence, the intention is to produce attraction and movement by using skewed perspective projections on statically positioned ambient displays.

This research suggests that embedding ambient displays that extend the visual depth into space can engage people to adhere to certain kinds of desired movement patterns.

Acknowledgments. The author would like to thank Katerina Alexiou, Sheep Dalton, Theodore Zamenopoulos and Jeff Johnson for their immense support throughout the study, Professor Paul Garthwaite for the statistical analysis and Zeta Kachri for helping, proofreading and editing the paper.

References

1. Wisneski, C., Ishii, H., Dahley, A., Gorbet, M., Brave, S., Ullmer, B., Yarin, P.: Ambient Displays: Turning Architectural Space into an Interface between People and Digital Information. In: Yuan, F., Konomi, S., Burkhardt, H.-J. (eds.) CoBuild 1998. LNCS, vol. 1370, p. 22. Springer, Heidelberg (1998)
2. O'Shea, C.: `http://www.chrisoshea.org/projects/out-of-bounds/`
3. Weiser, M.: The Computer of the 21st Century. Scientific American 265(3), 66–75 (1991)
4. Ishii, H., Ullmer, B.: Tangible Bits: Towards Seamless Interfaces between People, Bits and Atoms. In: Proc. of Conference on Human Factors in Computing Systems (CHI 1997), pp. 234–241. ACM Press, New York (1997)
5. Ishii, H., Wisneski, C., Brave, S., Dahley, A., Gorbet, M., Ullmer, B., Yarin, P.: AmbientROOM: Integrating Ambient Media with Architectural Space. In: Summary of Conference on Human Factors in Computing Systems, CHI 1998, pp. 173–174. ACM Press, New York (1998)
6. Porsche Museum (2009), `http://www.porsche.com/uk/aboutporsche/porschemuseum/`
7. ART+COM BMW Museum (2008), `http://www.artcom.de/index.php?lang=en&option=com_acprojects&id=64&Itemid=144&page=6`

8. Oosterhuis, K., Lénárd, I.: http://www.oosterhuis.nl/quickstart/index.php?id=492
9. Ataman, O., Rogers, J., Ilesanmi, A.: Redefining the Wall: Architecture, Materials and Macroelectronics. International Journal of Architectural Computing 04, 125–136 (2006)
10. Mathew, A., Taylor, J.: AuralScapes: Engaging Ludic Ambiguity in the Design of a Spatial System. In: Proc. of CHI 2008, pp. 2533–2542. ACM Press, New York (2008)
11. Tomitsch, M., Moere, A., Grechenig, T.: A Framework for Architecture as a Medium for Expression. In: Workshop on Pervasive Visual, Auditory and Alternative Modality Information Display, Pervasive (2008)
12. Röcker, C., Prante, T., Streitz, N.A., van Alphen, D.: Using Ambient Displays and Smart Artefacts to Support Community Interaction in Distributed Teams. In: Proc. of the OZCHI Conference, OZCHI 2004 (2004)
13. Fatah gen. Schieck, A., Briones, C., Mottram, C.: A Sense of Place and Pervasive Computing within the Urban Landscape. In: Proc. 6th International Space Syntax Symposium (2007)
14. Prante, T., Rocker, C., Streitz, N.A., Stenzel, R., Magerkurth, C., van Alphen, D., Plewe, D.A.: Hello. Wall– Beyond Ambient Displays. In: Dey, A.K., Schmidt, A., McCarthy, J.F. (eds.) UbiComp 2003. LNCS, vol. 2864, pp. 277–278. Springer, Heidelberg (2003)
15. Jafarinaimi, N., Forlizzi, J., Hurst, A., Zimmerman, J.: Breakaway: an Ambient Display Designed to Change Human Behavior. In: Proc. CHI 2005, pp. 1945–1948 (2005)
16. Consolvo, S., McDonald, D.W., Landay, J.A.: Theory-Driven Design Strategies for Technologies that Support Behavior Change in Everyday Life. In: Proc. CHI 2009, pp. 405–414 (2009)
17. Gibson, J.: The Ecological Approach to Visual Perception. Houghton Mifflin, Boston (1979)
18. Benedikt, M.: To Take Hold of Space: Isovist and Isovist Fields. Environment and Planning B 6, 47–65 (1979)
19. Hillier, B., Hanson, J.: The Social Logic of Space. Cambridge University Press, Cambridge (1984)
20. Hillier, B., Penn, A., Hanson, J., Grajewski, T., Xu, J.: Natural Movement: or, Configuration and Attraction in Urban Pedestrian Movement. Environment and Planning B: Planning and Design 20, 29–66 (1993)
21. Fatah gen. Schieck, A., Lopez de Vallejo, I., Penn, A.: Urban Space and Pervasive Systems. In: Proc. The Seventh International Conference on Ubiquitous Computing (2005)
22. Spiller, N.: Cyber_Reader: Critical Writings for the Digital Era. Phaidon Press, New York (2002)
23. Hillier, B.: Space is the machine. Syndicate Press, University of Cambridge, UK (2004)

Three User-Driven Innovation Methods for Co-creating Cloud Services

Ting-Ray Chang and Eija Kaasinen

VTT Technical Research Centre of Finland,
Tekniikankatu 1, P.O. Box 1300,
FI-33101 Tampere, Finland
{ting-jui.chang,eija.kaasinen}@vtt.fi

Abstract. The role of users in design is changing from one of passive research subjects to one of active co-designers. Users are the best experts of their everyday life experiences, making them great potential sources of innovation. User-driven innovation requires methods by which user ideas can be captured and worked on further with designers. In this paper, we describe our experiences of three different methods to co-create cloud services. Our aim was to innovate with users how open access to telecommunication data such as user location, user profile and usage logs could be utilised in cloud services. The user-driven innovation methods included focus group, crowdsourcing in an open web lab and face-to-face interaction in an open innovation showroom. We compare these three methods and identify the best usage possibilities for each. We propose guidance on selecting user-driven innovation methods based on the available resources and targeted results.

Keywords: User-Driven Innovation, Co-creation, Cloud Services, HCI research methods, Focus Group, Crowdsourcing, Open Web Lab (Owela), Open Innovation Showroom (Ihme).

1 Introduction

Today, human-centred design is an established practice for designing products and services so that the forthcoming users are represented in the design process. In a human-centred design process, user feedback is gathered throughout the iterative design process, and the design is refined accordingly. The process starts with identifying user needs by observing or interviewing users. In this phase, the decision to design a certain kind of product or service has already been made. To increase the user role in design and innovation, we should increasingly involve them in deciding what is needed and what kinds of products should be designed for them and with them. Kanstrup and Christiansen [1] describe this change as shifting the user role in design from a victim who needs support to a valuable source of inspiration. This change in the user role calls for user involvement in the innovation processes.

In user-driven innovation, users are no longer considered a reference group that sets the specifications for a system but a source of inspiration that can foster innovation [2]. User-driven design integrates user studies and idea generation [3]. A

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 66–83, 2011.

crucial factor for product and service development success is the performance in the early stages of the development process, that is a 'fuzzy front-end' when the targeted product or service is yet to be decided [4]. User involvement could be especially useful at this stage due to the high uncertainty and low formalization [5].

We have applied user-driven innovation to the case of future cloud services. Cloud computing refers to the paradigm shift from traditional installable applications to web-based software, where applications live on the Web as services. In cloud computing applications, hardware and system software are on remote data centres that can be located anywhere in the world [6]. In Cloud computing, resources can be purchased on demand from a virtually unlimited supply [7]. We have been studying a special case of cloud computing: Open Telco. Open Application Programming Interfaces (APIs) are a set of technologies that enable applications and services to interact with each other. This eases application design, as applications can utilise features and services provided by other service providers [8]. Open Telco is an open API framework for mobile networks. It gives external developers access to data and functionalities traditionally possessed by teleoperators. This enables open development of mash-up services that take advantage of network assets such as mobile payment capability, user profiles or location information [8].

In this paper, we describe and compare three approaches to user-driven innovation of Open Telco services. Our approaches were technology-driven rather than market pull, as we were dealing with radical improvements in functionality [9]. The user-driven design aim was to identify the meanings for the users of the various possible functionalities – the kinds of services that could be the forerunners of Open Telco. User-driven innovation activities were carried out using three methods. In focus group, users and designers met and ideated new services in a meeting room. In Open Web Lab, the ideation was taken to the web with social media. This made the ideation independent of time and place. The third study was carried out at an open innovation showroom in a real-world setting in which researchers and users interacted directly. The users could easily pop in when they had time, ideate with us, stay as long as they wanted and leave at their free will.

The paper is structured in the following way. Section 2 gives an overview of related research on user-driven innovation and co-creation. Section 3 describes the Open Telco case study. Section 4 describes how we applied the three user-driven innovation methods in practice and, finally, section 5 compares the methods and gives recommendations on how the methods can be applied.

2 Related Work

In marketing and human-computer interaction (HCI) research fields, the user role is changing from one of research subjects to one of co-creators. In marketing, one contributing factor has been the growing emphasis on services instead of products. A service provider has to establish, maintain and enhance the relationship with customers. This relationship-marketing approach emphasizes continuous interaction and co-creation with users [10]. User-driven design integrates user studies and idea generation and includes users as equal partners in the early phases of design [3]. In user-driven innovation, users are a source of inspiration that can foster innovation [2].

Participatory design focuses on designing social systems together with users and other stakeholders [11]. It has mainly been used to design work practices and work tools with different professionals. This is why participatory design is connected to a certain user community and usage context, whereas user-driven innovation may aim to innovate possible usages and usage contexts.

Hyvönen et al. [12] found direct contact between users and product developers to be an important element of user involvement. They emphasize the necessity of a common space where users and developers can meet. Sandström et al. [13] point out that functional elements of service experience are best met by services based on users' ideas. Ideas that improve the emotional elements of the service experience (like enjoyment and fun) can be difficult for users to generate spontaneously. This is why Sandström et al. recommend a combination of customer- and company-generated ideas for service innovation. In the case studies by Hyvönen et al. [12], focus groups that brought together users and developers turned out to be the most beneficial. The feedback from direct interaction was impressive and influencing.

The best known customer role in innovation is that of lead users introduced by von Hippel [14]. Lead users are characterized by their interest in the product, their frequent use of it or their activeness in information processing. Lead users have the ability to sense important market trends before the majority, which is why they may act as a 'needs-forecasting' laboratory. Alam [5] proposes that both lead users and ordinary users should be involved in the development process, as lead users are able to produce new ideas and their attractiveness can then be tested with ordinary users. An even more broadminded approach suggests that every user can offer something to the innovation process [15]. The only things that need to be determined are what kind of information is needed, what kinds of users or user groups are the best sources and what kind of interaction technique best enables the contribution.

In their study concerning open innovation communities such as CrowdSpirit and FellowForce, Antikainen et al. [16] differentiate between motivation to participate and motivation to collaborate. They identified, e.g., the following motivational factors for collaboration: interesting objectives, an open atmosphere, possibility to be influential, a sense of efficacy, having fun, a sense of community and rewards.

Design no longer just comes out of so-called 'black box' of designers' creative minds but involves different experts and users. Many different methods are used to address users' opinions throughout the development process, from requirement gathering to product evaluation. Different methods can be used to involve users in the design. Traditional methods such as focus groups, brain storming and thinking aloud have recently been extended with new methods such as cultural probes, co-design, online crowdsourcing and living labs.

Focus group studies bring together a group of users to discuss a predefined theme. The focus group session is usually organized in an in-house laboratory. Through discussions among participants, researchers can capture subjective insights from user perspectives [17].

While minor adjustments may be tuned according to the aims of the studies, focus groups usually take place in lab settings. One of the main reasons focus groups are commonly used in user ideation and evaluation is that they allow researchers to hear users' direct opinions easily and, in a discussion setting, users have to ground their opinions. The presented ideas can also be further developed in the discussion. Focus group results can be very helpful when making design decisions and in convincing engineers/marketing experts.

To shift the control from the designers' hands to the users' ones, several experimental methods have populated the HCI research world. With cultural probes [18], designers are inspired by collections of users' open and often random ideas from diaries, pictures or collages. In co-design, the user role is an equal design partner which enables the designers and users make design decisions together.

The transformation of the decision-making process in innovation from the designers' own judgment to the users' mass opinions is currently bringing in the best practices in user-driven innovation. Many innovative companies today put their ideas in crowdsourcing media for evaluation and further development before starting to realize them. Web and social media provide good tools for this. The idea of online crowdsourcing or crowd-wisdom has become popular not only because the ideation process is free from geographical constraints but also because it provides promising results from an open and less scrutinized environment [19].

Open Web Lab (Owela) is an online space for open innovation with users, customers, developers and other stakeholders [20]. Owela utilises social media features for participatory design and open innovation. It is a place where users, researchers and companies can meet and discuss ideas and critique. Owela provides tools for understanding user needs and experiences and for designing new products and services together with users [20].

This kind of online forum in which users are free to discuss and ideate according to their own schedule attracts many wild ideas and is easily accessible to people around the world. A similar environment is OpenIDEO, launched by IDEO, a well-known design and innovation consultancy. The OpenIDEO online platform is targeted at creative thinkers and uses 'challenges' as a method to discuss and solve problems. OpenIDEO currently focuses on social innovations.

The limitation of focus group studies often comes from the fact that the data from users are collected out of the real context and they therefore plant risks of validity when innovation is released into its context: the real world [21]. Outside the actual usage environment, the focus group may not think about all the issues related to usage situations. Online innovation spaces may have the same limitation.

Taking design activities into real-world environments changes the environmental setting of the user studies in HCI research from laboratories to the real world, such as living labs. Living labs are defined by the European Network of Living Labs (ENoLL) as real-life test and experimentation environments where users and producers co-create innovations [22]. According to ENoLL, living labs employ four main activities: Co-Creation, Exploration, Experimentation and Evaluation.

While living labs focus on permanent user communities, innovation showrooms aim to provide environments for short-term co-creation and innovation activities in real-life environments. The Ihme showroom (Figure 4) is an open innovation and co-design space located in a shopping centre. In this showroom, visitors can see and try applications of new technologies for themselves and are invited to ideate on how the technologies could be utilised.

Even if user-driven innovation has been studied quite a bit, the methods are not yet established as practices. We wanted to study how different user-driven innovation methods can address users' perspectives. We selected and performed studies with three user-driven innovation methods to compare them in terms of ease of organizing the innovation activity, user motivation to participate and quality of the outcome of the innovation. As the methods were tailored to investigate user insights according to the specific innovation challenge (Open Telco Cloud services), it is possible for the identified differences of the three methods to be case-specific. The results still give insights into the pros and cons of these methods for future studies.

3 Open Telco Case Study

Open Application Programming Interface (API) standards are already widely used on the Internet, and Open APIs are the core components of several Internet services. Open Telco is an open API framework for mobile networks. It takes the data and functionalities traditionally possessed by teleoperators to the cloud. External service providers can access the data via the cloud [8]. This enables external development of mash-up services that take advantage of network assets, such as call and message control, mobile payment capability, user profile, location information or presence [8,22].

Mash-up services combine data and functionality from several external sources. A mash-up approach has proven successful on the Internet application market, and similar phenomena can be expected with Open Telco. Open Telco has great potential to increase innovation of new mobile services, as high volumes enable reasonable pricing for location information, and standardized Open APIs enable low costs and innovative application development. Even niche services can be provided to small user groups [8].

Operators possess a large amount of customer history data, such as location and call logs, which could be utilised to profile target users, e.g., for advertisements [23]. Open Telco APIs can even be used in conjunction. For instance, the user can receive a call or a message launched by the event that (s)he arrives at a certain location. These kinds of trigger services can also be initiated by the presence or close location of a friend or a group [23].

In the Open Telco case, our aim was to identify the usage possibilities with most potential as well as the initial user requirements for Open Telco APIs. We used scenarios to illustrate the possibilities. The scenarios also served as a common language between the users and the designers. Our aim was to study broadly the different possibilities of the Open Telco infrastructure and then identify the applications with most potential from both the users' and service providers' points of view.

The Open Telco scenario work started with initial descriptions of different service possibilities identified in an earlier activity. We wrote initial scenarios based on these descriptions. We held workshops with project partners in which we refined the scenarios and made sure that they illustrated the key new service possibilities of Open Telco. The project group included technical, business and user experience expertise. After the project group had agreed on the scenarios, we started the user-driven innovation activities in which we assessed the scenarios with the users and let them propose new usage possibilities. The following value proposals were selected as the starting points for the user-driven innovation activities:

1. **Universal user profile** facilitates using the same profile for different services. This saves user efforts of updating separate profiles. The profile is always available via the cloud, even abroad.
2. **Peer recommendations** are received from the cloud. The recommendations can be selected based on the user profile and location, thus giving the user situationally relevant tips.
3. **Call and location history** can be collected in the cloud. The user can utilise the history data him/herself or the data can be used by service providers, for instance, to differentiate regular and occasional visitors.
4. **Mobile ticket with integrated services** is based on the possibility that the event organizer has contact with the ticket owners from the moment they buy the ticket. This facilitates different pre-during-after event services.
5. **Reverse-charge messages and calls** are based on the possibility that the user can ask the receiver to pay for a message or a call. This is useful in one-to-many messaging for instance.
6. **Machine-to-machine communication** services are based on a telecommunications facility embedded in different everyday objects.
7. **Future TV** could have many personal features similar to those of the Internet.

We illustrated value proposals 1-5 as PowerPoint clips with a main character named Matti to whom users could relate (Figure 1).

Different stories of Matti's life demonstrated what Open Telco value proposals could add to his life. Value proposals 6 and 7 were added as new value proposals after the workshop with the project team. With these value proposals, our aim was to generate user ideas for future TV services and user ideas for services based on machine-to-machine (M2M) communication. In the M2M and TV cases, our scenarios were very light as we just wanted to give the main idea to the users but did not want to turn their mindsets to a particular focus.

Fig. 1. An excerpt from a scenario illustration (universal travel profile)

4 Three User-Driven Innovation Methods in Practice

The aim of our user-driven innovation activities was to ideate with users of Open Telco services based on the value proposals presented in the previous section. We selected three different user-driven innovation research methods: focus group, online crowdsourcing in the open web lab and direct interaction in the open innovation showroom. The following sub-sections describe the three studies in more detail.

4.1 Focus Group Ideation

The main aim of the focus group study was to obtain quick feedback from potential users on the proposed Open Telco service ideas. Another aim was to ideate with the participants on the newly introduced value proposals of future TV and Machine-to-machine communications. The study included two parts:

1. Illustrated scenarios of value proposal 4 (Mobile ticket, Figure 2) and value proposal 5 (Reverse-charge Multimedia Message, MMS) were presented to the participants. The researchers mediated discussions in which the participants gave feedback and ideas.
2. Value proposal 6 (Machine-to-machine communication, M2M) and value proposal 7 (Future TV) were presented to the participants by giving simple real-life examples. The participants were then asked to ideate in pairs on near future possibilities for Open Telco services.

Fig. 2. Illustration of a mobile ticket scenario

Eight in-house scientists (four female and four male) who work in the area of ICT for Health participated in the focus group study. One researcher chaired the focus group and another observed and took notes. The scenario of value proposal 4 was presented to the participants with the PowerPoint illustrations. After the presentation, the following themes were discussed:

- What did you like or dislike about the presented service?
- Can you suggest anything that could be added to the service?
- Could you foresee this service on the market? Would you recommend it to your friends? Why or why not?

After assessing the scenario of value proposal 4, the scenario of value proposal 5 was assessed in the same way.

After the two scenario evaluations, the participants were asked to ideate about future TV and machine-to-machine communication (M2M). The introduction to M2M ideation was as follows: '*Imagine everyday objects/machines like waste bins, cars and TVs talking to each other. Which ones would they be and how?'* The participants were asked to discuss their ideas in pairs and to write the ideas down on post-it notes. After the ideation, each pair presented their favourite ideas and the ideas were discussed in the group.

The TV ideation followed the same procedure as the M2M did. As an example of future TV services, a Yahoo!-connected TV widgets demo was presented. The participants were asked: '*Imagine what else you could do on your TV. What other web-related functions would you like to have on your TV?'*

The entire focus group session lasted one hour. The data collected included notes taken by the researchers and ideation post-its from the participants. The user feedback on the scenarios illustrating value proposal 4 (Mobile ticket) and value proposal 5 (Reverse-charge MMS) were satisfactory in terms of acceptance and novelty value. The ideations for value proposal 6 (Machine-to-machine communication) and value proposal 7 (Future TV) were short but fruitful. Several participants were excited about sharing their ideas by telling a full story. The researchers gained concrete ideas on how Open Telco services applied to future TV and how machine-to-machine communication could benefit users. The participants also gave ideas on how the proposed applications could be presented to attract users.

4.2 Online Crowdsourcing in the Open Web Lab (Owela)

The second study was carried out as online crowdsourcing in an open web lab (Owela). The ideation was based on the scenarios illustrating value proposals 1 (Universal user profile), 2 (Peer recommendations) and 3 (Call and location history). Another aim was to ideate usage possibilities for the new value proposals 6 (Machine-to-machine communication) and 7 (Future TV). The study included two parts:

1. Online users were briefed with texts and illustrated scenarios of value proposals 1, 2 and 3. This was followed by open discussions under different sectors of the Owela online forum, led by initial questions set up by the researchers. Participants could give feedback and ideas on the illustrated value proposals.
2. Value proposal 6 (Machine-to-machine communication) and value proposal 7 (Future TV) were illustrated by giving simple real-life examples with texts and images. Online users were free to use the 'add an idea' button as well as to comment on others' ideas and the initial ideas supplied by the researchers.

Fig. 3. Screenshot of the Owela online ideation page for machine-to-machine communication

The online studies lasted a month in total. During the first two weeks, new users were invited from university email newsletters, online international student groups and Facebook to discuss the first three value proposals. Forty-six new users joined Owela during the first two weeks to participate in our study. During the following two weeks, existing Owela users were also invited and the two new value proposals were added (Future TV and Machine-to-machine communication). The participants came from 11 countries in Europe, North America and Asia. All the users used English as the communication language. Owela online ideations and evaluations involved a total of 84 users. There were 48 males, 33 females and 3 unknowns. The ages of the participants ranged from 18 to 68. Two researchers participated in moderating the discussions. The users had to register in the project discussion area to start. The users did not have to reveal their identities, but their online identities were displayed beside their comments. Many of the comments turned into discussions between the researchers and the users, and among the users.

The participants were briefed with an introduction page and were then able to proceed freely to the value proposals they wanted to see and comment on. Five tabs (Figure 3) took the users to the ideation pages where each value proposal was introduced using texts and illustrations, followed by questions from the researchers to encourage the users to leave comments and ideas. As an example, Figure 3 presents the ideation page for machine-to-machine communication.

During the ideation, the researchers stimulated the discussion with additional questions and comments. The overall Owela study produced a total of 221 user comments in a blog style, sectioned by topics with canonical orders. The results were analysed and the ideas grouped according to application areas and themes. Many innovative and wild ideas were produced as well as direct and passionate discussions on personal experiences and life stories. Some discussions ended with a general consensus on preferences of service ideas and acceptance. Other discussions ended with various directions of personal preferences, many with good reasons. The discussions illustrated the user ideas well and grounded the users' opinions. It was challenging to analyse the outcomes of the discussions systematically however.

4.3 Direct Interaction at Ihme, the Open Innovation Showroom

The main aim of the third study, direct interaction at an open innovation showroom (Ihme), was to present our ideas in a real-world setting and to interact directly with ordinary users as a new approach to user-driven innovation. The Ihme innovation showroom and co-creation space was located in a shopping centre (Figure 4). There were also other technology demonstrations such as a virtual travel experience and a 3D TV at the Ihme showroom. People who came to the Ihme showroom were free to have a look at any or all of the demos.

Similar to the Owela study, it was carried out with the scenarios illustrating value proposal 1 (Universal user profile), 2 (Peer recommendations) and 3 (Call and location history). The ideation process of value proposals 6 and 7 in this study was meant to be less structured and freer. The study was carried out in two parts:

1. The participants were briefed with illustrated scenarios of value proposals 1, 2 and 3. This was followed by an open discussion with the researcher and led by initial questions set up by the researcher.
2. Value proposals 6 and 7 (Machine-to-machine communication and Future TV) were illustrated as posters. Users were able to add their comments and ideas as post-it notes and study the notes left by other users.

Fig. 4. Ihme innovation showroom

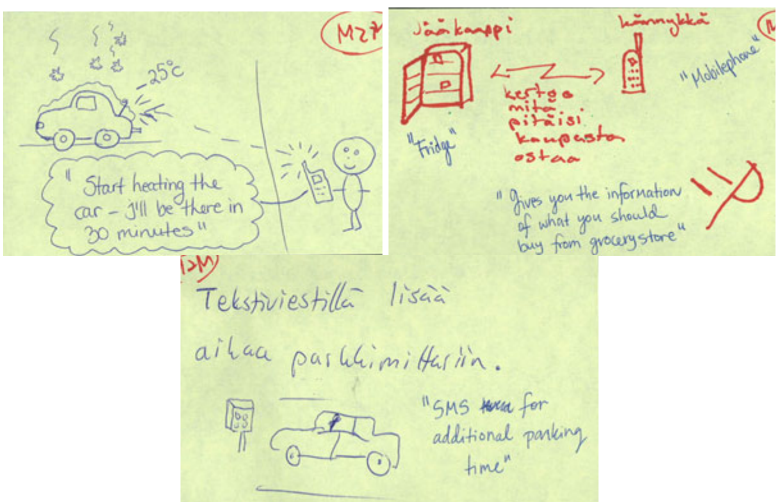


Fig. 5. Post-it notes from user ideations on machine-to-machine communication

The first period of the studies ran for nine days. The researchers invited ordinary people from the shopping centre to participate in the studies and interviewed a total of 26 people. As the interviews were conducted in an informal way, user demographics were not asked. Compared with the previous two studies, many of the participants in this study were less technology-oriented. The recruitment prioritized English, though Finnish researchers were also available. The participants were first briefed with a set of PowerPoint illustrations of the three first scenarios for a total of two to three minutes. Each participant was then interviewed by a researcher with a list of twelve predefined questions that covered personal preferences and acceptance of each value proposal, as well as ideas for additional services. The interviews were semi-structured and interactive. Many participants therefore used their personal experiences to express their opinions. The interviews lasted ten to twenty minutes.

The next period of this user study focused on the ideation of value proposals 6 and 7. Two posters illustrated the ideas. Users were invited to leave their ideas as post-it notes on the posters. The users could see others' comments and were able to comment freely, add a new idea or continue with the available ideas. For each comment, the participants received a lottery ticket to win cinema tickets. Each poster was displayed at the Ihme showroom for a week and the post-it notes were collected at the end of the week. Researchers were available so that users could discuss the ideas with them. In total, 13 post-it ideas were collected for the 'future TV' poster and 5 for the 'machine-to-machine communication'.

Each ideation session with individual users lasted from 30 minutes to about 1 hour, depending on the interaction and discussion between the researcher and the participants. The data collected included notes taken by the researcher and ideation post-its from the participants. The user feedback on the scenarios of value proposals 1-3 was satisfactory in terms of acceptance and novelty value. Due to the semi-structured discussion, the results were quantitative and structural. The ideations for value proposals 6 and 7 were not large in terms of quantity but interesting in terms of expressions. Figure 5 illustrates some of the post-its left by the participants.

It was challenging to tempt people who were passing by to participate in the study for a moderately long time. Many of the participants had come to the shopping centre with their families in their spare time. It turned out to be important to present the value proposals in such a way that ordinary, less technology-oriented users could quickly understand them. The flexibility of the schedule, according to each individual's available time, was also important. Direct interaction with users was fruitful and produced concrete ideas. In a real-world setting, the users were more interested in commenting on familiar topics such as future TV. Machine-to-machine communication was difficult to understand and several users commented that they did not need to know about it.

5 Comparison of the Approaches

An overview of the setups and results of the three user-driven innovation studies is presented in Table 1. The focus group is clearly a quick method that can be carried out in one or two hours, whereas Ihme and, especially, Owela usually require longer periods of time. An Ihme study could be carried out in one day, but the session for each participant would then have to be planned quite compact, and a good strategy to recruit users would be necessary. A focus group works well if the targeted user group is very specific, e.g., a certain work community. Owela is very useful for international studies as it overcomes the restrictions of time and place. Ihme is at its best when ordinary users are the target group. In the following sub-sections, we will analyse the differences in the methods with regard to the setup of the study, the user motivation to participate and the quality of the results.

Table 1. Comparison of the three co-creation methods

	Focus group	**Owela**	**Ihme**
Length of the study	1 hour	1 month	3 weeks
Language	English	English	English/Finnish
No. of users involved	8 participants, 4 F + 4M	84 online users, 49 M + 35 F	26 users
Designer:user ratio	2:8	2:84	1:1
Procedure	Briefing→Ideation →Discussion	Briefing→ Ideation+Discussion	Briefing→ Ideation+Discussion
Study duration	1 hour	4 weeks	20~40 min per user, 3wks total
Location	Finland	International	Finland
Value proposals	4 (4, 5 and 6, 7)	5 (1, 2, 3 and 6, 7)	5 (1, 2, 3 and 6, 7)
Environment	Laboratory	Real world	Real world
Illustration material	PowerPoint, video clips of scenarios	Screen shots of scenarios	PowerPoint, video clips of scenarios
Media	Face to face	Online	Face to face
Data gathering	Notes and post-its	Online forum	Interview notes and post-its
No. of comments	63	221	252
Data from Scen. 1-3	Acceptance comments + 5 new* ideas	Acceptance comments + 20 new* ideas	Questionnaire data (19 questions)
Different categories of ideas – future TV	4	12	13
Different categories of ideas – M2M	5	12	5
Group/individual discussions	Group	Group	Individual
User participation, scheduled versus free	Scheduled	Free	Free

5.1 Setting Up the Studies

The illustration material for all three studies was a semi-automated PowerPoint slide show of scenarios in which the main character, Matti, was involved in stories that introduced different value proposals in everyday situations. The same material can basically be used with all three methods, with some subtle differences. In the Owela web lab, there is no possibility of oral introduction by the researcher. Therefore the illustrations of the proposed services must be very clear and easy to understand. The illustrations also have to be quite compact as the participants are not motivated to study long presentations. In Ihme, the material has to be designed to include elements that will tempt potential participants to come and have a closer look. The posters worked well for this purpose. In the focus group and Ihme, the slide shows worked, but for Owela the scenarios had to be shortened and, as the researchers were not always present, separate questions for each scenario had to be included. Effort was also made in the Owela study to ensure that the value proposal in the scenario was understandable without a human introducer.

In our case, the Owela and Ihme spaces with related human resources were already available. It was therefore quite easy to set up and perform the studies. The setting up of such research environments from scratch could involve an intensive workload.

5.2 Motivating Users

The studies in the real-world environments, open web lab (Owela) and open innovation showroom (Ihme) attracted participants from more diverse backgrounds than the focus group as a laboratory setting to which participants were selected and invited for the session. The participants in Owela were found to be more technology oriented. Ihme in the shopping centre environment attracted more ordinary users.

In Owela, users were very motivated when they were interested in the topics. Some users revisited the site often to follow up new discussions. It is challenging, however, to raise the discussion if the topic is not interesting enough to the participants. Topics such as privacy issues or acceptance raised the most elaborate discussions. The Owela participants felt free to discuss how they felt without face-to-face confrontations. It can be concluded that a good Owela theme is one that raises feelings for or against.

In the focus group, the participants focused for a shorter period of time from understanding the theme to finishing the discussion. They may have been motivated but may not have had enough time to become inspired and ideate based on their real-life experiences.

In Ihme, the users' interests and focuses often depended on their schedule, as their visit to the shopping centre may have included prior plans and goals to achieve. The Ihme participation was not agreed in advance and the participants needed to be on a more flexible schedule to be involved. The ideation theme also has to be interesting to the participants and, in addition, some tempting elements should be presented to attract a closer look, e.g., an interesting illustration on a poster. Innovation showrooms are best suited to the ideation of tangible experiences with prototypes or simulated services. The Ihme innovation methods need to be developed to also allow short visits while still offering the users the possibility to participate.

The compensations for Ihme and Owela were based on a lottery whereas in the focus group, the compensation was just refreshments. With all three methods, small compensations were attractive ways to motivate participants.

5.3 Data Quality

The three qualitative studies produced many insights into the users in a 'thick descriptive way', which is more fruitful in such ideation activity than quantitative results. The three methods can also be compared quantitatively, as shown in Table 1. Ideation data are presented in terms of quantity (number of comments) as well as quality (different categories of ideas produced). In Ihme, the researcher discussed ideas with the participants and took notes at the same time instead of recording. This made note-taking quite challenging. Owela's online group discussions collected most user ideas both in terms of quantity and quality. Owela comments gave a wide and in-depth range of user insights from daily experience sharing.

While all three studies produced rich, qualitative data, we found the data from the Ihme studies to be easily quantifiable, as the users' answers were documented following a set of questions and answered individually. Comparatively, it took more time and analysis to understand and summarize the data collected from Owela. The main difficulty was that the material was in discussion threads and the same user could have commented several times throughout the thread.

Owela provided a platform that was free of time and space constrains for users to ideate and discuss. Ihme provided a co-creating space where researchers could really see and feel what users thought about cloud services (e.g., when a user took his phone and started acting out a scenario related to his idea). In the direct face-to-face interaction in Ihme and the focus group, we were able to understand the participants as individuals better than in Owela in which the participants were not present in person. In Ihme, the users were also given the chance to understand the technical possibilities by asking questions in direct interaction. In all three studies, the designers and researchers benefited from the various degrees of communication with the users. In the focus group, the researchers had investigator roles. In online crowdsourcing (Owela), the researchers mostly had observer roles and partly discussion partner roles. In the open innovation showroom (Ihme), the researchers had a more equal role to that of the users in a co-creating activity.

In the Owela web environment, the participants were almost anonymous (only an online identity was displayed beside the user's comment). Anonymity may have made the participants feel free to express themselves at their own pace and style. Consensus or influenced discussions may occur in Owela discussions (they did not happen in our study). It is then the researchers' duty to observe carefully or mediate the flow of discussion to stay on track. The permanent user community in Owela that was invited during the second round can be seen as a kind of lead user group. They had participated in different ideation activities and were more familiar with being involved in the discussions. The ideation diversity of our study was fruitful, possibly due to the mixture of lead and new users.

The comments and discussions in a focus group tend to carry a chance of redundancy. Our focus group participants sometimes commented easily by agreeing with previous speakers or were otherwise influenced by them.

In Ihme, each participant gave individual answers. Compared with the other two studies as group discussions, the comments were not influenced by other participants. From Ihme, it was therefore possible to report an overview of the participants' opinions in a quantitative way, e.g., 23 out of 26 are interested in trying cloud services. The ideations in Ihme may be less stimulated without continuous group discussions however. In the future, the Ihme method need to be developed so that they also facilitate the user-user dialogue and wider group discussions.

In the focus group, we had more short answers, such as 'yes I agree' and 'yes I agree with the previous comments', which did not provide in-depth information. In Owela we had large amounts of text in the comments, and some comments were actually life stories. Even if Owela comments were time-consuming to analyse, many useful insights came out of the data. In Ihme, the researchers were able to interact with the users, listen to the users' comments and then ask further questions straight away or discuss the proposed idea. In Ihme, the participants based their comments on thinking about situations in which they would use the service, as the following comments related to privacy show:

> '*It [peer recommendation services] feels like being tracked when you are on the road and everyone will make the same choices.*'
>
> '*If it [universal profile] follows me everywhere, it is like a chip in my dog...*'

In Owela, the participants based their comments on both personal and general use, as the following comments show:

> *'There are going to be serious privacy issues though, so this would definitely need to be an opt-in service so people don't get upset.'*
> *'I think I would be a bit worried about who gets access to my information and possible misuse issues.'*
> *'It seems obvious to me that this information cannot possibly be kept private. It already bugs me that my car can be tracked as I cross bridges or toll points.'*

In both Owela and the focus group, ideas were developed further in discussion. In Owela, however, the participants could comment and ideate according to their own schedule. In Owela the comments were often long and well analysed. Figure 6 illustrates an excerpt from an Owela ideation chain related to Future TV.

Participant (P) 1: 'TV scrapbook – what if I could copy and paste, e.g., recipes from a cooking show to my video cookbook (maintained in the cloud)?'

(con-)P 2 'I would love to get them on my mobile phone… would save me time at the store too if I could check for missing items in the recipe …'

(con-)P 3: 'I would like my TV to collect a scrapbook from news and documents of topics I am interested in. For instance I save the word "environment" in its memory and it follows up all news and documents that deal with the environment and it collects a scrapbook on environment!'

(con-)P 4: 'Definitely a good idea! The semantic TV! I think as well as transmitting pictures, sound and subtitles, each frame/second/minute (whatever is practical) should have information that tells the TV/computer what the content in the picture means and potentially connects to other information. If something interesting appears in the picture, hit the Info button and the TV/computer gives a list of information relating to what is in the picture...'

Fig. 6. An excerpt from an Owela ideation chain on Future TV

With all three methods, we aimed to gain a firm idea of our users of the services by trying to understand the users' perspectives – 'walking a mile in the users' shoes'. With the development of user-driven innovation research methods, we are taken from observing from afar to a close-up understanding of our users. As a lesson learnt from collecting and analysing research data, user values and acceptances were easy and fruitful to discuss in the Owela online ideation. In the focus group, the users gave short answers that only provided a consensus that they were concerned about privacy issues in the scenarios related to the profile and call and location history. In Owela, the users were very comfortable discussing their personal experiences openly and providing solutions that made them feel more secure. In Owela, it is also possible to see how many people have similar or different opinions in the group discussions but not in a consensus way to produce only one conclusion. In Ihme, the researchers and users discussed issues in a side-by-side way and shared experiences.

If the data from the three studies were to be labelled in some way, it would be: the focus group was the most efficient way of producing quick ideas; Owela produced the most creative ideas; and Ihme produced the ideas that are closest to real-life (tangible) experiences.

6 Conclusions

Focus group studies have been one of the best established user research methods in the HCI field. The focus groups are also well-suited to user-driven innovation. A well-defined service proposal can receive quick feedback in a focus group within a short time. Based on our studies, there are also other alternatives for focus groups depending on the targeted users and results.

Online crowdsourcing is a good method for reaching a large, international group of users. Registered Owela participants can be utilised as a lead user group as they are a kind of ideation professional. In the web environments, the users felt free to talk about their personal experiences, such as dislikes and fears. Personal preferences and issues such as privacy and ethics could be discussed thoroughly and the outcome could be very well grounded. Innovation by the crowd is a powerful method in which ideas can be fomented. The ideation theme has to be interesting enough to motivate the users to participate and continue their participation however. The ideation theme has to be such that it can be understood quickly without excess documentation. Online discussion material may be hard to analyse afterwards as the material is in different discussion threads. If quantitative data are needed, separate voting activities can be organized during the discussion.

The co-creation of tangible experiences, services and products could benefit most from demonstrations in an open innovation showroom such as Ihme. In Ihme, the ideation theme has to be such that it tempts passers-by to have a closer look. As in Ihme, the sessions were not scheduled beforehand and the schedule could become a challenge. Ideally, each user should be able to devote as much (or as little) time to the ideation as (s)he happens to have. This will require new methods for Ihme studies. The Ihme innovation showroom facilitated direct user-designer interaction in the actual usage context. As suggested by earlier studies, this dialogue can be very powerful as it not only foments ideas but also makes the designer understand the user's world. In our study, users only exchanged and commented on ideas by other users in the latter part of the study with post-it notes on the idea poster. Methods that would facilitate wider communication with other users are a future need for Ihme.

Acknowledgement. Our research was carried out as a part of *Tivit*, the Strategic Centre for Science, Technology and Innovation in the Field of ICT (www.tivit.fi), the *Cloud Software Program* (www.cloudsoftwareprogram.org). The work is partly funded by *Tekes* (the Finnish Funding Agency for Technology and Innovation, www.tekes.fi). We would like to thank our research partners in the Cloud Program, especially *TeliaSonera* and *Aalto University* with whom the Open Telco research was carried out. We would also like to thank our colleagues at *VTT* who assisted with the Owela and Ihme studies. We are most grateful to all the users who contributed in the co-creation activities.

References

[1] Kanstrup, A.M., Christiansen, E.: Selecting and evoking innovators: Combining democracy and creativity. In: Proceedings of the 4th Nordic Conference on Human-computer Interaction: Changing Roles (2006)

[2] Holmquist, L.: User-driven innovation in the future applications lab. In: Conference on Human Factors in Computing Systems (2004)
[3] Botero, A., Vihavainen, S., Karhu, K.: From closed to open to what? An exploration on community innovation principles. In: MindTrek 2009, Tampere (2009)
[4] Khurana, A., Rosenthal, S.R.: Towards holistic 'front-end' in new product development. The Journal of Product Innovation Management 15(1), 57–74 (1998)
[5] Alam, I.: Removing the fuzziness from the fuzzy front-end of service innovations through customer interactions. Industrial Marketing Management 35, 468–480 (2006)
[6] Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Stoica, I.: A view of cloud computing. Communications of the ACM 53(4), 50–58 (2010)
[7] Miller, M.: Cloud computing: Web-based applications that change the way you work and collaborate online. Que Publishing Company (2008)
[8] Suikkola, V.: Open exposure of telco capabilities – identification of critical success factors for location-based services in Open Telco. In: 6th International Conference Wireless and Mobile Communications (ICWMC), pp. 202–208 (2010)
[9] Verganti, R.: Design, meanings, and radical innovation: A metamodel and a research agenda. Journal of Product Innovation Management 25(5), 436–456 (2008)
[10] Grönroos, C.: Service management and marketing: A customer relationship management approach. John Wiley & Sons Inc., Chichester (2000)
[11] Muller, M.J., Kuhn, S.: Participatory design. Communications of the ACM 36(6), 24–28 (1993)
[12] Hyvönen, K., Heiskanen, E., Repo, P., Saastamoinen, M.: Kuluttajat tuotekehittäjinä: Haasteita ja mahdollisuuksia (Consumers as product developers: challenges and possibilities). In: Kuluttajat kehittäjinä, pp. 31–48. Kuluttajatutkimuslaitos, Helsinki (2007)
[13] Sandstrom, S., Magnusson, P., Kristensson, P.: Increased understanding of service experiences through involving users in service development. European Journal of Innovation Management 12(2), 243–256 (2009)
[14] Von Hippel, E.: Lead users: A source of novel product concepts. Management Science 32(7), 791–805 (1986)
[15] Heiskanen, E., Hyvönen, K., Repo, P., Saastamoinen, M.: Käyttäjät tuotekehittäjinä (Users as product developers). Teknologiakatsaus 216. Tekes (2007)
[16] Antikainen, M., Vaataja, H.: Rewarding in open innovation communities – how to motivate members. International Journal of Entrepreneurship and Innovation Management 11(4), 440–456 (2010)
[17] Nielsen, J.: Usability Engineering. Academic Press, Boston (1993)
[18] Gaver, B., Dunne, T., Pacenti, T.: Culture Probes. Interactions 6(1), 21–29 (1999)
[19] DeVun, L.: Looking at how crowds produce and present art. Wired News, November 19 (2009), http://www.wired.com/techbiz/media/news/2007/07/crowd_captain?currentPage=all
[20] Näkki, P., Antikainen, M., Virtanen, T.: Participatory design in an open web laboratory owela'. In: Proc.CHI 2008. ACM, New York (2008)
[21] Genise, P.: Usability Evaluation: Methods and Techniques. Version 2.0. University of Texas (2002)
[22] European Network of Living Labs, http://www.openlivinglabs.eu/
[23] Raivio, Y., Luukkainen, S., Juntunen, A.: Open Telco: A new business potential. In: Proceedings of the 6th International Conference on Mobile Technology, Application & Systems, pp. 1–6

Designing for the Secondary User Experience

Ole Andreas Alsos and Dag Svanæs

Department of Computer and Information Science,
Norwegian University of Science and Technology, Trondheim, Norway
{Ole.Andreas.Alsos,Dag.Svanes}@idi.ntnu.no

Abstract. Computer systems are first and foremost designed for primary users. The needs of the other types of users, such as secondary users, are often overlooked. These users are not interacting with the system directly but are yet affected by it. This study is based on empirical findings from two usability evaluations in a realistic hospital setting with physicians and patient actors. We have found that also secondary users, such as patients, have a kind of user experience during the primary user's interaction with the system. We conclude from this that designers and developers should also address the need of secondary users and include them in the design and evaluation process. This means designing devices or GUIs that (1) support non-verbal communication, (2) provide feedback to the secondary users, (3) use their language and representation, and (4) is tailored for the secondary user. Sometimes a focus on the secondary user implies that the designer must deal with conflicting needs between the primary and the secondary users.

Keywords: User experience, UX, end user, secondary user, secondary user experience, patient experience, clinical simulation, usability evaluation.

1 Introduction

Most *user interfaces* are designed for single users. This is true even for the user interfaces of collaborative applications, such as Facebook and Skype (although the applications are collaborative). However, as systems are being used in collaborative settings, the use of computing systems will increasingly involve people beside the primary user. Sometimes other people indirectly become users of the user interface without even interacting with it. One such situation is face-to-face interactions where one person, the *primary user*, is using an information system while interacting with another person face-to-face, i.e. the *secondary user*. It could be a physician talking with a patient in a medical consultation while interacting with an electronic patient record system, or a clerk serving a customer while looking up information in a product directory (e.g. in banks, travel agencies, or shops). In such situations one could expect that both persons are affected directly or indirectly by the user interface of the information system; both the primary user, who is using the system directly, and the secondary user, who relies on the primary user to obtain information from the system and is affected by the primary user's experiences with the system.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 84–91, 2011.

The needs of the secondary users are rarely in the minds of the system designers and developers of information systems, and are often neglected [1]. While the *user experience* (UX) is considered important when designing for the primary user, there is no equivalent concept in the HCI literature for the secondary users.

In this paper we seek to extend focus on user experience to go beyond the traditional “first person user” perspective and to include secondary users. We do this by defining the *secondary user experience*. Then, by drawing on two different usability evaluation studies of mobile devices used in a realistic hospital setting, we identify some ways in which the design of the user interface can affect the secondary user experience. Finally, we provide preliminary guidelines for designing for the secondary user experience in a hospital setting.

2 Secondary User Experience

Eason [2] identifies three categories of users; (1) *primary users* who are frequent hands-on users of the system; (2) *secondary user*s who are occasional users or use the system through an intermediary, and (3) *tertiary users,* who are affected by the introduction of the system or influence its purchase. Other researchers have defined other comparable end-user variations [3, 4].

It is the authors’ impression that designers and developers to some extent are aware of the peripheral end-user groups, i.e. those who do not directly interact with the systems, but that few actually design for them. They mainly take the direct/primary user into account when designing system. Thorough work is often done in identifying stakeholders in the requirements process, but end-users are mainly considered those who directly interact with the system (e.g. [1]).

The current usability definition, ISO 9241-11:1998 [5], does not distinguish between different end-users. However, a number of HCI publications refer to a 1997 draft of the definition. This version contained a reference to people that do not directly interact with the system: “[Satisfaction is] the comfort and acceptability of the work system to its users *and other people affected by its use*” (our emphasis). However, this reference to “other people” was omitted in the final version of the usability standard (see [6] for an example).

In many use situations, especially from client-service relations and health care, the primary users of information systems are engaged in face-to-face interaction with customers or patients. They may for example be a client making deposits in a bank, a globetrotter booking a flight in a travel agency, or a customer buying products in a shop. The way we see it, these types of secondary users are recognized by the following characteristics:

- They are interacting with the primary user, who interacts with the system.
- They are not (or in little extent) interacting directly with the system themselves.
- They rely on the primary user to obtain information from the system.
- They are influenced by the primary user’s experiences with the system (e.g. effectiveness, efficiency, satisfaction, etc.)

2.1 User Experience and Co-experience

Researchers and practitioners have different views on what the *user experience* (UX) is [7]. As defined in ISO 9421-210 [8], it is “a person’s perceptions and responses that result from the use or anticipated use of a product, system or service". UX is here associated with the primary user.

Others have proposed comparable definitions, such as Law et al., [7], who propose that “UX focuses on interaction between a person and something that has a user interface”. The study excludes face-to-face interaction between people, unless a user interface is involved in the interaction. For this kind of interactions, *Co-Experience* provides a better explanation; it is the user experience created in social interaction with the presence of a system or product [9]. For use situations as described in the introduction, UX and co-experience relates as follows: While UX concerns the primary user, co-experience relates to both the primary and secondary user (Figure 1).

2.2 Secondary User Experience

We have found few examples of secondary user experiences in literature. An exception is provided by Montague [10], who claims that patients have a user experience of technology used by care providers even if they are passive users of it. During childbirth, the health care providers use technology to monitor the health condition of the unborn child. By interviewing mothers about the technology used on them, Montague found that when technology worked well, it created positive experiences and increased the patients’ connection with their babies. When technology did not work well, or when care providers could not get technologies to work properly, negative experiences occurred.

We have found no existing definitions of the user experience of secondary users. Therefore we define it as follows:

> *The secondary user experience of a system is the part of the overall experience of the secondary user that can be attributed to (1) the primary user’s interaction with the system, or (2) the secondary user’s interaction with the system with the primary user as an intermediary.*

Secondary UX is different from UX and co-experience in that it only relates to the experiences of the secondary user (see Figure 1).

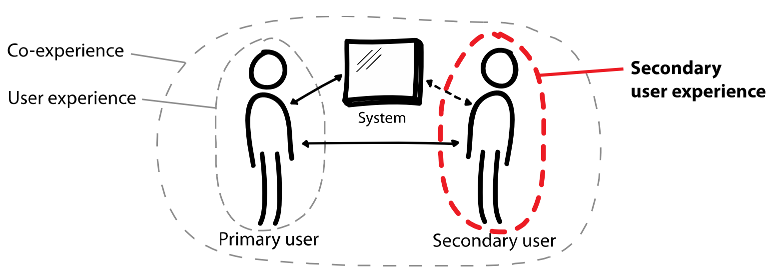


Fig. 1. *UX* relates to the primary user, *Co-experience* relates to the shared experience by the system of both users, and *secondary UX* relates to the secondary user. The solid arrows indicate interaction, while the dotted arrow indicates indirect interaction.

3 Methods

The ward round is a situation with both a primary user (the doctor) and a secondary user of a system (the patient). Mobile electronic patient record systems for bedside usage are primarily designed with the doctor in mind. Few have considered the effects on the patient. Therefore, we have retrospectively analyzed the data and findings from two previously conducted usability evaluation studies of prototype mobile systems for hospitals [11, 12]. Both evaluations were conducted in a simulated hospital environment with multiple persons involved; real physicians consulting patient actors lying in the hospital beds.

Using handhelds and patient terminals together. In the first evaluation, we explored several ways of allowing doctors to use handheld devices together with bedside mounted patient terminals for viewing x-ray images in collaboration with the patient (Figure 2). Five pairs consisting of a physician and a patient each performed eight ward rounds using different variants of the x-ray image viewer. In total, 40 ward rounds where performed. For each evaluation, the physician and the patient were interviewed simultaneously about aspects concerning the user experience of the system. The prototypes and research approach are previously described in [11].

Handheld medication system. In the second study we explored interaction techniques for a handheld medication system; one paper based and three mobile patient record systems (Figure 3). 14 physicians conducting in total 56 ward rounds. Physicians and the patients were interviewed separately about aspects concerning the user experience. The study and prototypes are previously described in [12].

With assistance from a physician and a nurse, much effort was used designing the patient scenarios, tasks and environment for both evaluations as close to their real working environment as possible. The patient actors were carefully instructed how to behave and what to say, to make the behavior as realistic as possible. Senior sociologists with experience from health care studies, as well as facilitators and participating physicians and patients considered the scenarios played out in the evaluations to be realistic instances of hospital work.

The video recordings from the evaluations were analyzed using methods from video-based studies of human interaction with technology [13], while the interview data was analyzed qualitatively using methods influenced by grounded theory.

Fig. 2. Using handheld devices together with bedside mounted patient terminals

Fig. 3. Comparing interaction techniques for handheld medication system

4 Results

When replacing the non-digital solutions, such as paper based records and x-rays images with mobile computing technology, we found a number of issues related to the user experience for the primary and secondary users:

The primary users preferred digital over non-digital. Although the physicians in general were confident in using the paper chart (study 2), they preferred using the mobile computing device. A number of functions and attributes, such as pocket size, error prevention and undo mechanisms, contributed positively to the UX.

Secondary users had strong opinions about unused system. In the first study the patients, as passive users of the system, had strong opinions about the different versions of the x-ray viewer, even if the patient's part of the system was identical (or nearly identical) for all versions. The main difference was on the mobile device, which was only used by the physician. However, the physicians' usage of the different variants of the system had large effects, both on how smoothly they were able to present new x-ray images on the patient terminal, and how both the physicians and patients rated them afterwards.

Mobile devices hampered the nonverbal communication. The physical form factor of the paper chart allowed the physician to use it as a channel for nonverbal communication. For example, some doctors signaled that the consultation was ending by closing the chart, or invited the patients to speak by tilting the chart towards them. This was harder with the PDA, and was considered negative by the patients.

Mobile devices reduced the doctor-patient dialogue. The user interface of the mobile device increased legibility and allowed the physicians to undo and minimize medication errors. On the other hand, the user interface gave poor information overview and had unfamiliar interaction techniques. This required much of the physicians' attention, and according to the patients it hampered the doctor-patient dialogue and reduced their satisfaction of the consultation.

Good for the doctors was (sometimes) bad for the patients. In most of the design solutions of the first study, the doctor controlled the patient terminal through a mobile device. While this was seen as a major benefit from the perspective of the physicians, who could hide private information on the mobile device and display public information on the patient terminal, it was perceived as negative from the perspective of the patients. They did not understand what was going on and did not like that information was hidden for them.

Good for the patients was (sometimes) bad for the doctors. In other design solutions the physicians controlled the system directly through the patient terminal. Unintentionally it also allowed the patient to partly interact with the terminal. For the patient this was perceived as an improvement. For the physicians, however, it became harder to control the system as they had to bend over the patient to use it.

User interface complexity confused patients. In some versions of the first study, the controls for changing information content were present on the patient terminal. The increased complexity of the GUI confused some patients. They rather preferred the versions where these controls were moved onto the physician's mobile device.

Spoken language was influenced by GUI. In the second study, a pause symbol (as used in music and video players) was used on the mobile device to temporarily cease medical treatment without removing it from the medication list. This lead the physicians to use words like "pausing this drug" rather than the more incomprehensible term "cessation", which was commonly used when using the paper chart. While the first term was obvious for the patients, the latter was a foreign word.

5 Discussion

The findings in this study demonstrate that secondary users have a kind of user experience of an information system that is used by primary users. Further it shows that designers sometimes face tradeoffs between the primary and secondary UX.

5.1 User Experience is Relevant for Secondary Users

The studies demonstrated, not surprisingly, that technology had an impact on the primary user (i.e. physician), who was directly interacting with the system. Further, our observations showed that the primary users' interaction with the system also had an impact on the secondary users. For example, the patients had some sort of user experience; they had strong perceptions and responses about the system, although they had not used the systems directly themselves.

The positive correlation between patient satisfaction on health outcome has long been established [14]. When patients report that they are satisfied or dissatisfied because of the physician's interaction with the system (i.e. their experiences as a secondary user), we can assume that it has some impact on the overall patient satisfaction. This was for example seen when the complexity of the GUI confused patients.

5.2 Trade-Off between Primary and Secondary User Experience

The analysis indicates that designers are faced with tradeoffs between the needs of the primary and secondary users. When the user experience was improved for the physicians, it had in some cases negative effects for the patients, e.g. the ability to hide information on the mobile device. In addition, we found that when the secondary user experience was improved, it sometimes created new problems for the physicians, e.g. reducing the ergonomics when interacting with the system. Consequently, aspects of the user experience for the primary user can have negative consequences for the secondary user. In a similar manner, improving the user experience for the secondary user can have negative consequences for the primary user.

How should one deal with the potential trade-offs between the design of the primary and secondary user experience? One cannot design the primary user experience first and then the secondary user experience. It may lead to a suboptimal solution for the latter. Both user experiences must therefore be designed together. Yet, the designers must often prioritize. In the hospital it could be a bad priority to propose designs that threat patient safety just to make the patient experience slightly better.

5.3 First Steps towards Design Guidelines

Accommodating the needs of the secondary users is important. In the context of a ward round with a physician and patients, a positive secondary user experience can have positive effect on the doctor patient dialogue, which is important for the treatment and care of the patients [14]. In the context of business, for example a travel agent serving a potential traveler, or a checker handling a customer, improving the secondary user experience can have positive effect on the overall customer experience. This often means satisfied returning customers and increased revenue [15]. Drawing on our findings, we suggest four preliminary design guidelines that we find relevant for information systems involving a secondary user experience. Research in other domains and with other technologies will be needed to make the list more complete:

- *Give system feedback to the secondary user.* By increasing the action transparency (i.e. increase the visibility of actions) or providing alternative system feedback to the secondary user, one can improve the secondary user experience.
- *Support non-verbal communication.* The quality of the non-verbal aspects of face-to-face communication has a strong impact on the secondary user experience. Our findings indicate that the system can hinder this communication, especially when the system occupies the hands or hides the face of the primary user. Therefore, the physical form factor of the system needs to support nonverbal communication.
- *Use the language and representation of the secondary user.* By presenting the information for the primary user in the language of the secondary user, the primary user can be guided to use simpler terms and communicate on the same level as them, i.e. physicians use terms like “blood sugar level” instead of “glucose”. This can make it easier for the secondary user to understand.
- *Provide a GUI and/or device tailored for the secondary user.* If feasible and necessary, an additional device/GUI with information tailored for the secondary user should be provided. This will give the secondary users a version of the information where unnecessary complexity and irrelevant information is removed.

In addition to the above suggestions, we suggest that (1) the perspectives of the secondary user are included throughout the design process, and that (2) usability evaluations include both the primary and secondary users *together*.

5.4 Limitations and Future Work

It is important to note that the evaluation studies presented here were not explicitly designed for the purpose of illuminating the secondary user experience. The data was analyzed retrospectively. Moreover, we recognize the limitations regarding the use of “role-plays” as one of our primary data sources.

The design guidelines are very preliminary and their validity is limited to ward round scenarios with mobile technology. We do believe, however, that other domains with secondary users may benefit from similar studies to improve the secondary user experience. Future work should further investigate the concept of secondary user experience, as well as the generalizability outside the presented case.

6 Conclusion

The needs of secondary users, who only use an information system occasionally or through an intermediary, are often neglected. Designers and developers must address the need of secondary users and include them in the design and evaluation process of the system. This means designing devices or GUIs that (1) support non-verbal communication, (2) provide feedback to the secondary users, (3) use their language and representations, and (4) is tailored for the secondary user. Sometimes this implies that the designers deal with conflicting needs between primary and secondary users.

References

1. Sharp, H., Finkelstein, A., Galal, G.: Stakeholder identification in the requirements engineering process. In: Proceedings DEXA 1999. IEEE, Los Alamitos (1999)
2. Eason, K.: Information Technology and Organisational Change. Taylor and Francis, Abington (1987)
3. Faulkner, X.: Usability Engineering. Palgrave Macmillan, Basingstoke (2000)
4. Ågerfalk, P.J.: Who is the user in user-centered design? In: HCII. Springer, Heidelberg (2001)
5. ISO 9241-11:1998. Ergonomics of human system interaction - Part 11: Guidance on usability specification and measures. International Organization for Standardization
6. O'Grady, M.J., O'Hare, G.M.P., Sas, C.: Mobile agents for mobile tourists: a user evaluation of Gulliver's Genie. Interacting with Computers 17 (2005)
7. Law, E.L.-C., Roto, V., Hassenzahl, M., et al.: Understanding, scoping and defining user experience: a survey approach. In: Proceedings of CHI 2009. ACM, New York (2009)
8. ISO 9241-210:2009. Ergonomics of human system interaction - Part 210: Human-centred design for interactive systems. International Organization for Standardization
9. Battarbee, K.: Defining Co-Experience. In: Proceedings of DPPI 2003. ACM, New York (2003)
10. Montague, E.N.H.: Understanding Patient User Experience in Obstetric Work Systems. Health (San Francisco), 70–77 (2009)
11. Alsos, O.A., Svanæs, D.: Interaction techniques for using handhelds and PCs together in a clinical setting. In: Proceedings of NordiCHI 2006, pp. 125–134. ACM, New York (2006)
12. Alsos, O.A., Dabelow, B., Faxvaag, A.: Doctors' Concerns of PDAs in the Ward Round Situation — Lessons from a Formative Simulation Study. Methods Inf. Med. 48(6) (2009)
13. Heath, C., Hindmarsh, J., Luff, P.: Video in qualitative research. SAGE Pub. Inc., Thousand Oaks (2010)
14. Ong, L.M.L., de Haes, J.C.J.M., Hoos, A.M., et al.: Doctor-patient communication: A review of the literature. Social Science & Medicine 40(7), 903–918 (1995)
15. Peppers, D., Rogers, M.: Return on customer: creating maximum value from your scarcest resource. Currency (2005)

Engaging Visitors in Museums with Technology: Scales for the Measurement of Visitor and Multimedia Guide Experience

Mohd Kamal Othman, Helen Petrie, and Christopher Power

Human Computer Interaction Research Group,
Department of Computer Science,
University of York,
Heslington East YO10 5GH UK
{kamal,Helen.Petrie,cpower}@cs.york.ac.uk

Abstract. Mobile technologies such as multimedia guides (MMGs) are now an important part of the visitor experience in museums and other cultural spaces. We report the development of two scales to measuring visitors' museum experiences: the Museum Experience Scale (MES) and the Multimedia Guide Scale (MGS); these quantitative measures can helpfully complement qualitative information about visitor experience. A standard psychometric methodology was used in the development of these scales: from a large set of potentially relevant statements, 57 were chosen and 255 people rated a museum experience (102 of whom had used a multimedia guide). A Principal Components analysis yielded a four factor solution for the MES (Engagement, Knowledge/Learning, Meaningful Experience and Emotional Connection) and a three factor solution for the MMGS (General Usability, Learnability and Control, Quality of Interaction). Comparing respondents who used a MMG during their museum visit with those who did not, there was a significant difference on the Engagement component of the MES, with respondents who used a MMG being significantly more engaged. The other components of the MES did not show significant differences.

Keywords: Museums, cultural spaces, user experience, multimedia guides, audio guides.

1 Introduction

Museums and other cultural institutions such as art galleries, historic houses, and archeological sites, referred to in this paper as *cultural spaces*, have been using various technologies to improve their visitors' experiences for nearly 60 years. The Stedlijk Museum in Amsterdam was the first museum to use a handheld guide in their exhibitions in 1952 [1]. It took nearly a decade before other cultural spaces followed that example, with the American Museum of National History adopting the "Sound Trek" audio guide in 1961. In addition, a Sony Walkman type system was created for the famous "Treasures of Tutankhamun" tour in the late 1970's, whilst the Louvre museum introduced the first random access guide in 1993.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 92–99, 2011.

Emerging technologies such as smart phones and tablet computers are now further changing the way technologies are used in cultural spaces. In addition, the use of technology in cultural spaces is now not limited to audio commentary, but may provide diverse content types such as images, video and multimedia. A recent study found that 57% of cultural spaces surveyed in North America, Asia and Europe have adopted multimedia guides [2].

It is important for cultural spaces to embrace new technologies to engage and stimulate their visitors in exhibitions. The use of these technologies should not be regarded as replacement of the curated tour or more traditional means to disseminate information, but instead as further ways to connect and engage visitors with objects, collections and exhibits. Wasserman argues that the use of mobile technology is more than an information-distribution platform, and that it should instead connect visitors with each other, with the institution playing an important role of bringing people together through shared experiences [3].

Further, Pekarik argues that in order to have diverse and exciting cultural spaces in the future there must be investigation through rigorous methods on how to increase the range of satisfying experiences had by visitors [4]. Clearly, technology can play a major role in helping to create those experiences. However, if technology is not developed and deployed carefully it is likely that technology could detract from the visitors' experience. Therefore, it is important to develop methods and measures for determining the nature, valence and size of the effects that technology has on visitors' experiences in cultural spaces.

There are many examples measuring different aspects of user experience that are useful for studying technology in the cultural spaces domain. The *acceptance* of technology by users, a multi-dimensional concept comprised of perceived ease of use and perceived usefulness of the technology, was originally proposed by Davis and studied in many different domains [5]. These factors are also discussed in work on cognitive absorption [6]. There are examples of measuring immersion, the feeling of being pulled into and becoming lost in the interaction with a piece of technology [7, 8], and qualitative work in examining the engagement of visitors with interactive exhibits [9]. Finally, there is *flow* [10], the concept of optimal experience where a user experiences feelings of satisfaction and achievement when the experience is complete. This, too, has been examined in many different contexts [11, 12, 13]. However, none of these measures, as defined in previous work, have been developed specifically for the cultural spaces domain. In this paper, we build on previous work on user experience to create measurement scales for this domain to quantify the effect that the use technology has on visitors.

This study was designed to investigate the effects of multimedia guides on the experiences of visitors to cultural spaces. We will take both a qualitative and quantitative approach to this question. We are developing standard questionnaires to measure both visitors' overall experience, particularly the engagement, with the exhibition (the Museum Experience Scale, henceforth MES) and the usefulness and usability of multimedia guides (the Multimedia Guide Scale, henceforth MMGS). We have presented initial findings on the MES [14] that produced four factor structure, here we present the final form and the MMGS.

This paper will outline the full development of the two scales, and results of an initial use of the MES, as well as plans for further development of the research.

2 Method

2.1 Scale Development

The development of the two questionnaires followed standard psychometric scale development [14, 15]. Initially, a large set of topics, ideas and statements were collected by reviewing relevant previous studies and papers [5, 8, 9, 16, 17, 18, 19] and materials developed by the UK Council for Museums, Libraries and Archives. The Generic Learning Outcomes model developed by the MLA [18] was particularly useful in developing the range of statements. This process resulted in 152 potential statements for the scales[1]. Three evaluators then used a consensus process to reduce the number of statements by grouping them into themes and removing similar or overlapping statements. This resulted in 57 statements: 37 for the MES and 20 for the MMGS. The items were presented in the scales as Likert items from 1 meaning "strongly disagree" to 5 meaning "strongly agree".

2.2 Procedure

The initial versions of the two scales were presented online using QuestionPro[2] survey software. The study was widely publicized via numerous email lists and an advertisement on Facebook. Publicity asked for people who had visited a museum in the past six months, with or without a multimedia guide. To encourage participation, a prize draw for Amazon gift vouchers was offered to all participants.

All participants completed the initial version of the MES and of the MMGS if they had used a multimedia or audio guide on their museum visit. In addition, they also completed a short questionnaire to gather information about their museum visit (which museum, how long the visit lasted, how many people in the party etc), as well as standard demographic information.

2.3 Participants

255 participants completed the scales. 96 were male, 175 were female. Participants came from very diverse backgrounds (e.g. country of residence, education or work background, age). This should help to ensure a robust scale. 102 respondents had used multimedia guide during their museum visit, 153 participants had not.

3 Results

For each scale, a principal components analysis was performed on the ratings of the statements to extract the *components* or groups of questions that elicit similar responses from participants.

[1] The set of potential statements is available at: these are available at `www.cs.york.ac.uk/hci/docs/inital_items.pdf`

[2] `http://www.questionpro.com/`

Table 1. The 4 components of the Museum Experience Scale (MES) and their factor loadings

Engagement		**Knowledge/Learning**	
I enjoyed visiting the exhibition	0.69	The information provided about the exhibits was clear	0.64
I felt engaged with the exhibition	0.69	I could make sense of most of the things and saw and did at the exhibition	0.57
My visit to the exhibition was very interesting	0.68	I liked graphics associated with the exhibition	0.52
I felt I was experiencing the exhibition, rather than just visiting it	0.65	My visit enriched my knowledge and understanding about specific exhibits	0.52
My visit to the exhibition was inspiring	0.56	I discovered new information from the exhibits	0.43
Meaningful Experience		**Emotional Connection**	
During my visit I was able to reflect on the significance of the exhibits and their meaning	0.74	The exhibition enabled me to reminisce about my past	0.55
During my visit, I put a lot of effort into thinking about the exhibition	0.53	My sense of being in the exhibition was stronger than my sense of being in the real world (reversed relationship)	0.52
Seeing rare exhibits gave me a sense of wonder about the exhibition	0.50	I was overwhelmed with the aesthetic/beauty aspect of the exhibits	0.47
After visiting the exhibition, I was still interested to know more about the topic of the exhibition	0.43	I wanted to own exhibits like those that I saw in the exhibition	0.45
Seeing real exhibits of importance was most satisfying aspect of my visit to the exhibition	0.43	I felt connected with the exhibits	0.45

The MES produced four components:

- *Engagement* with the exhibitions and exhibits
- *Knowledge/Learning* gained from the exhibition and exhibits
- *Meaningful Experiences* from the interaction with the exhibitions/exhibits and/or other visitors
- *Emotional Connection* with the exhibits/exhibitions

Factor loadings on the top questions for each component are shown in the Table 1, below. A *factor loading* is a measure of how strongly each statement relates to the overall component (1.0 = perfect relationship to 0.0 = no relationship at all, only statements with factor loading over 0.43 are listed). From this analysis, a final selection of 20 questions (5 questions for each component) for the MES was made. For example, if two of more similar statements were in the same category, the statement with the higher factor loading was selected.

Table 2. The 3 components on the Multimedia Guide Scale (MMGS) and their factor loadings

General Usability		**Learnability and Control**	
I will use an multimedia guide again when I visit an exhibition (negative correlation)	0.76	I felt I was in control of the multimedia guide	0.78
The multimedia guide was a distraction	0.74	Learning to operate the multimedia guide was easy	0.74
The information given by the multimedia guide was too lengthy	0.73	Using the multimedia guide did not require much training	0.70
It was difficult to determine where I was in the exhibition with the multimedia guide	0.68	The controls of the multimedia guide were difficult to understand (negative correlation)	0.64
The multimedia guide helped me to navigate around the exhibition (negative correlation)	0.67	The multimedia guide presented information in an understandable manner	0.54
Using the multimedia guide enhanced my exhibition visit (negative correlation)	0.65	I found it difficult to read the text on the screen of the multimedia guide (negative)	0.53
The multimedia guide was complicated to use	0.51		
It was difficult to select the option I wanted with the multimedia guide	0.51		
Quality of interaction with the Guide			
The multimedia guide clearly provided feedback about my actions	0.72		
It was clear to me when the multimedia guide was taking the initiative to offer me information and when I needed to ask it for information	0.54		
I became unaware that I was even using any controls on the multimedia guide	0.48		

The principal components analysis for the MMGS yielded three components:

- *General usability* of the multimedia guide, for example whether the functionality of the guide is appropriate, whether it is easy to use
- *Learnability and control,* whether the guide is easy to learn to use, whether the user felt in control, and whether the information presented in a meaningful ways
- *Quality of interaction with the guide*, this is often considered part of usability or user experience, but interestingly in this scale, the aspects concerning interaction with and feedback from the guide form a separate component

Table 2 shows the statements that relate to each component and their factor loadings. The questions that are labelled "negative correlation" mean that high ratings on those questions are associated with low scores on the scale.

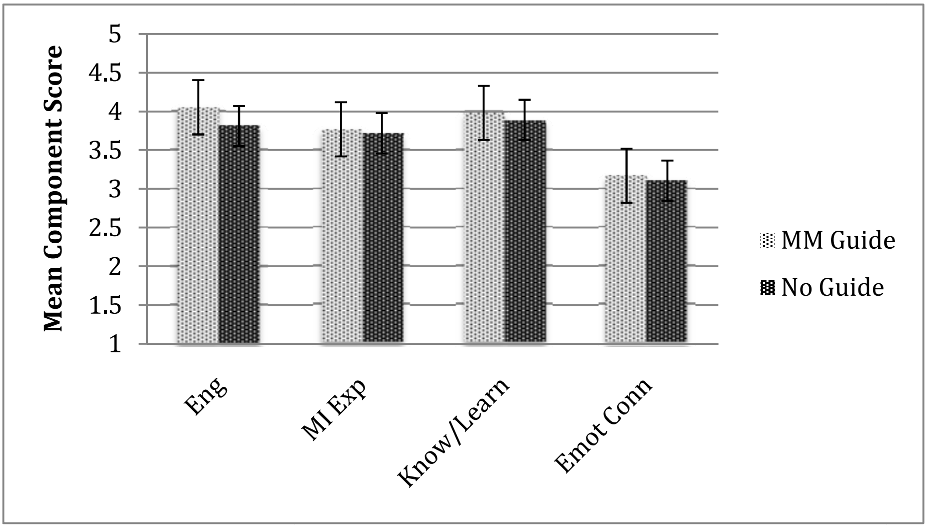


Fig. 1. Mean scores on the four components for multimedia guide and non-guide users

As an initial use of the MES, we compared the experience of participants who had made their museum visit with a multimedia guide with the experience of those who had made their visit without a multimedia guide. There was an overall significant difference in scores between two groups of participants (Analysis of variance $F_{1, 253} = 3.66$, $p < 0.05$). There was also a significant difference between the four factors ($F = 149.50$, df = 3, 759, $p < 0.001$). There was no interaction between the group and factor variables.

Figure 1 shows the mean scores on the four MES components for multimedia guide and non-guide users. This shows that scores on all four components were higher (i.e. more positive) with a guide, although that difference was only significantly higher on the Engagement component, with multimedia guide users being significantly more engaged than non-guide users (Tukey's HSD $p < 0.05$). The lack of significant difference on the other three components is interesting in itself. This shows that the use of a multimedia guide has no effect, positive or negative on the Knowledge and Learning, Meaningful Experience, or Emotional Connection about an exhibition for museum visitors.

4 Discussion and Conclusions

This study has developed two scales for use in the evaluation of visitors' experiences of museums and other cultural spaces. We are not suggesting that these are the only instruments that are needed in the evaluation of such visitor experiences, but that they can be a useful part of an "evaluation toolkit" available to personnel responsible for evaluations. Particularly as mobile technologies such as multimedia guides become an integral part of the museum and cultural space experience, it is important to have tools available to assist in their evaluation.

The Museum Experience Scale (MES) produced four components: Engagement, Meaningful Experience, Knowledge/Learning and Emotional Connection. These are slightly different components from those we found from the initial analysis of data [14], but the current structure is based on a much larger and more robust sample of respondents. It is interesting that only the Engagement component produced a significant difference between multimedia guide users and non-users. Thus use of a multimedia guide appears to enhance engagement and does not detract from a meaningful experience of emotional connection with the exhibition.

Clearly the sense and level of engagement with exhibitions and exhibits in museums varies between visitors. Prior knowledge, motivation, interest, technology, and time spent in the exhibition may influence engagement. The findings from our study show that introducing technologies such as multimedia guides are achieving their aim, to make the museum experience more engaging for visitors. This finding supports previous research that the use of handheld devices such as multimedia guides is more appealing to visitors than more conventional and traditional ways of presenting information [18].

Although this study was a useful contribution in quantifying user experience in museums with and without multimedia guides, it had several limitations. In particular, respondents were asked about museum experiences that might be six months old, which was not ideal. To address this and several other issues, we now planning a study that will use the two scales in an on-site study of visitor experience with an archeological exhibition. This will allow us to collect data immediately after the visitor experience and to concentrate on the experience in one particular museum. Later research will be able to investigate and compare other types of museum and cultural spaces and other types of technology. In our next study, we will use two different forms of multimedia guide to investigate how differences in the presentation of the guide can affect visitors' experience of a museum or cultural space. We will also compare this with other measures of visitor experience such as short interviews.

References

1. Tallon, L.: Introduction: Mobile, Digital and Personal. In: Tallon, L., Walker, K. (eds.) Digital Technologies and the Museum Experience: Handheld Guides and Other Media. Altamira Press (2008)
2. Petrie, M., Tallon, L.: The iPhone effect?: Comparing Visitors' and Museum Professionals Evolving Expectations Mobile Interpretation Tools. In: Trant, J., Bearman, D. (eds.) Museums and the Web 2010. Archives and Museums Informatics, Toronto (2010)
3. Wasserman, S.: Beyond Information: Ritual, Relationship and Re-encounter through Mobile Connectivity. Curator: The Museum Journal 54, 11–24 (2011)
4. Pekarik, A.J.: The Long Horizon: The Shared Value of Museums. Curator: The Museum Journal 54, 75–78 (2011)
5. Davis, F.D., Bagozzi, R.P., Warshaw, P.R.: User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. Management Science 35, 982–1003 (1989)

6. Agarwal, R., Karahana, E.: Time Flies When You're Having Fun: Cognitive Absorption and Beliefs about Information Technology Usage. MIS Quarterly 24, 465–694 (2000)
7. Cheng, K., Cairns, P.A.: Behaviour, Realism and Immersion in Games. In: CHI 2005 Extended Abstracts on Human Factors in Computing Systems. ACM Press, Portland (2005)
8. Haywood, N., Cairns, P.: Engagement with an Interactive Museum Exhibit. In: People and Computers XIX - The Bigger Picture. Springer, Heidelberg (2005)
9. Jennett, C., Cox, A.L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., Walton, A.: Measuring and Defining the Experience of Immersion in Games. International Journal of Human Computer Studies 66, 641–661 (2008)
10. Csikszentmihalyi, M.: Flow: The Psychology of Optimal Experience. Harper and Row, New York (1990)
11. Chen, H., Wigand, R.T., Nilan, M.: Optimal Flow Experience in Web Navigation: Effective Utilization and Management of Emerging Information Technologies. In: Ninth Information Resources Management Association Conference. Idea Group Publishing, USA (1998)
12. Rettie, R.: An Exploration of Flow during Internet Use. Internet Research: Electronic Networking Applications and Policy 11, 103–113 (2001)
13. Gaggioli, A., Bassi, M., Delle Fave, A.: Quality of Experience in Virtual Environments. In: Riva, G., Davide, F., Ijsselstijn, W.A. (eds.) Being There: Concepts, Effect and Measurement in User Presence in Synthetic Environments. IOS Press, Amsterdam (2003)
14. Othman, K., Petrie, H., Power, C.: Understanding visitors' experiences with multimedia guides in cultural spaces. In: Aljaset, A., et al. (eds.) Proceedings of Transforming Culture in the Digital Age. University of Tartu, Tartu (2010)
15. Anastasi, A., Urbina, S.: Psychological Testing, 7th edn. Prentice Hall, Upper Saddle River (1997)
16. DeVellis, R.F.: Scale development: theory and applications, 2nd edn. Sage Publications, London (2003)
17. Naismith, L., Sharples, M., Ting, J.: Evaluation of CAERUS: A Context Aware Mobile Guide. In: mLearn 2005, Cape Town (2005)
18. Naismith, L., Smith, P.M.: Using Mobile Technologies for Multimedia Tours in a Traditional Museum Setting. In: mLearn 2006, Canada (2006)
19. MLA. Inspiring Learning for All, Museum and Libraries Association, `http://www.inspiringlearningforall.gov.uk`
20. Pekarik, A.J., Doering, Z., Karns, D.A.: Exploring Satisfying Experiences in Museums. Curator: The Museum Journal 42, 117–129 (1999)
21. Boehner, K., Gay, G., Larkin, C.: Drawing Evaluation into Design for Mobile Computing: A Case study of the Renwick's Gallery's Hand Held Education Project. International Journal of Digital Libraries 5, 219–230 (2005)

An Image of Electricity: Towards an Understanding of How People Perceive Electricity

Yoram Chisik

Madeira Interactive Technologies Institute (M-ITI) - University of Madeira
Campus da Penteada, 9020-105 Funchal, Portugal
ychisik@m-iti.org

Abstract. Although an enormous amount of research effort has been devoted to understanding people's energy consumption habits, visualizing their consumption and finding ways of motivating them towards more sustainable behaviours we are still in the dark with regards to people's basic perception of electricity, their concept of what electricity is and their notion of the consumption rates of various electrical devices. In this study we have employed a sketching methodology to elicit people's basic mental image of what electricity is, how they conceive of the electrical infrastructure in their home and which devices they think represent the largest drain on their wallets. Preliminary analysis of the results show that people do not have a clear mental model of electricity and tend to associate the size of the device and the duration of use with higher rates of consumption regardless of the type of device, the type of use it is put to and its actual consumption level.

1 Introduction

Electricity is the staple of modern life. Electrical devices from air-conditioners and light bulbs to vacuum cleaners and washing machines enable us to escape the harshness of the weather, turn night into day, and make what was once dirty clean again. Electricity powers our communication networks and fuels our production facilities, it brings to life the alarm clocks and coffee makers that wake us up in the early morning and energizes the displays that keep us busy during the day and entertained at night.

Electricity is also hidden from view as the current powering a device flows sight unseen from the generator or grid. Unless we fall victim to an incompetent electrician, or play with live wires because we have lost, or are yet to acquire our senses, most of us will never come into direct contact with live electricity, although there is always the possibility of getting struck by lightening or getting shocked by a static charge regardless of our sense or sensibility.

Thus electricity remains largely out of sight and out of mind while we regale in the use of the multitude of electrical devices that make up the fabric of the modern world. It is the invisible nature of electricity that lies at the root of some of the environmental problems currently being faced by the world's population, as behind most electrical devices lies a wall socket, and behind the wall socket lie a maze of wires and cables which more often than not lead to the belching smokestack of a power generating station.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 100–117, 2011.

As we try to move towards a more sustainable future, curtailing the consumption of electricity and its associated emissions has turned out to be a particularly wicked problem. By using the term *wicked problem* I refer to both the standard dictionary definition of a highly serious and distressing problem and the term coined by Rittel and Webber [26] of a complex problem for which there is neither a definitive formulation nor a definitive solution.

The extent of the wickedness of the problem is revealed when one studies the literature on the subject. Not only are there issues with the sitting of power stations, the fuels and other natural resources used to power them and the complexities of managing the network of cables and wires through which the electricity flows, research into the use and consumption of energy and electricity have revealed that human nature and proclivity add a set of new dimensions to the problem.

To begin with, Pierce, Schiano and Paulos [23] have shown that people are highly resistant to change and value their convenience when using appliances, i.e. they may not forsake or delay their use of a device due to environmental considerations. While Chetty, et al [4] have observed that even persons intent on shutting down their computers during periods of inactivity stopped doing so due to the inconvenience associated with the long waits caused by the prolonged start-up and shutdown times of the computer.

To compound the problem even further consumption patterns both in terms of purchasing specific appliances and in terms of use are highly context specific, and are influenced by a host of variables ranging from personal preferences and economic ability (both past and present) to weather patterns and social norms [1], [19] and are often conducted without conscious consideration [17], [22].

HCI researchers have sought to bring the "unconscious aspects of [electricity consumption] to conscious awareness, thereby making them available for conscious choice" [27]. Pierce, Odom and Blevis [21], Froehlich, Findlater and Landay [11] and Pierce and Paulos [22] describe the various eco-visualization devices that have been constructed in order to make visible the flow of electrical energy and to highlight the amount of electricity consumed by a household. He and Greenberg [13], Mankoff, et al [16], and Foster, et al [9] describe the ways in which social networking systems and computer mediated communication have been used as means of supporting and motivating energy conservation and sustainable consumption.

However these approaches in addition to having to overcome the difficulties outlined above also have to contend with economic reality. Studies have shown that the financial incentives associated with smart-meters and eco-visualization devices are often too small to motivate behavioural change. Smart-meters and eco-visualization devices typically bring about a 5-15% reduction on average [6], [28] and since there are no long term studies of the use of smart meters and eco-visualizations we do not know whether these savings can be sustained over time.

Dunlap [8], Bostrom, et al [2], Read, et al [25] conducted a series of studies aimed at finding what people know and understand about global warming. What they found was that although respondents showed concern over the problem their knowledge and understanding was highly limited. The respondents tended to confuse the processes involved (often having difficulty in differentiating between causes and actions) or offer highly literal interpretations. Dunlap [8] also found that these perceptions did not consistently vary across social strata and that most people were willing to admit their lack of knowledge and understanding.

Thus much depends upon people's preferences and perceptions yet despite a tremendous amount of research work we are still in the dark with regards to how people perceive electricity. A search of the literature focused specifically on electricity and perception yielded a single reference [14] and even that dealt with people's perceptions of the aesthetic elements surrounding the placement of electrical poles and not with electricity itself.

If we are to make significant inroads into the subject of electricity consumption we first need to unsheathe the cloak of invisibility from electricity and understand how people conceive electricity and electrical consumption as opposed to examining the ways in which they use electrical devices and their perceptions of the impact of that usage on the environment which has been the primary focus of most electricity consumption studies to date. This study is our first step in building such an understanding.

2 Research Questions

Electricity is both a natural phenomenon and commercially available product. As a natural phenomenon electricity is most often experienced either as the physical sensation of a minor shock caused by a static electricity charge when reaching for a door knob or a railing, or as a visual sensation such as a lightning spectacle during a storm.

As a commercial product electricity is available in numerous forms of delivery and in different "strengths". Electrical utilities offer their customers a steady stream of electricity via a purpose built network for a fee. Battery manufacturers offer electricity in packaged form and generator manufacturers sell devices that produce electricity from sources as varied as gasoline and solar rays and in capacities ranging from the minute to the gargantuan.

The prevalence of commercial electricity is such that the vast majority of the world population consumes electricity in one form or another throughout the day. Those fortunate enough to live in the more developed parts of the world enjoy a constant stream of electricity in their homes and workplaces and power their mobile devices with a host of batteries. Those living in lesser-developed or remote sites use generators to either supplement or to provide the entirety of their electricity needs. Regardless of the level of development, isolation and socio-economic status one would be hard-pressed to find a location entirely devoid of some form of electricity or electrical consumption.

In all cases the electricity delivered and its quantity will be hidden from view. Unlike other forms of energy such as cow dung, wood, coal, petrol, gas or other consumables such as water, fruits or vegetables commercial electricity has no visible or felt properties. One cannot spill a bit of electricity from a cable like one would be able to do with water or petrol from a pipe, or feel whether a battery is full or empty by sensing its weight, as one would do with a water bottle or a gas canister. Electricity is an invisible resource that can neither be seen nor heard and its presence can only be inferred from the workings of the devices that are dependent on it. Indeed most of the time people only think of electricity during a power failure when the lights and the rest of the appliances on which their life depends stop working.

As a result the mental model of electricity, i.e the perceptual image a person stores in his head and on which he bases his reasoning and decision will be based on mediated/indirect experience of using electrical devices as opposed to the direct experiences that shape our mental models of physical space and physical properties. For example we form mental models of landscapes by physically traversing the streets of a city, the aisles of a supermarket or the corridor of a friend's apartment, noting the layout of the environment, e.g. 3rd door on the right and/or specific memorable features (landmarks) such as the fountain or the duck at the corner of the road.

RQ1 – What are people's perceptions of electricity?

Given the fact that electricity is virtually universal in both its natural and commercial forms, that practically no man alive today has known a world entirely devoid of commercial electricity and that we can only experience commercial electricity indirectly through the agency of some device it would be interesting as a starting point to find out what is the first thing that comes up in people's mind when they think of electricity.

Would people think of one of the electrical devices they use on a regular basis, would they invoke images of one of the infrastructure elements through which electricity is delivered to their homes, offices and mobiles such as electricity poles, plugs, cords and batteries, would an image of a lightning storm flash in their minds or would they think of something else entirely.

RQ2 – How do people conceptualize the electrical infrastructure within their homes?

As noted above people's relationship with electricity is an indirect one. The power of electricity is utilized through the use of electrical devices and the absence of electricity is often only felt when these devices cease to work

The average home contains a complex array of wiring, access points and devices within and between its walls. Some of the wiring is hidden from view, encapsulated into the walls, while other is visible and in plain sight. A few devices such as the refrigerator are autonomous in nature, they are permanently connected to the electrical system, run all the time and are completely self regulated[1]. Other devices such as washing machines, microwave ovens and televisions are fixed in place and constantly hooked up to the electricity supply via a wall socket either due to their size or because they are often used but only operate and therefore draw power when switched on by the consumer (either directly or via a timer setting)[2].

In addition a whole host of devices ranging from small kitchen appliances through hair dryers and electric drills to mobile devices such as cell phones and media players are stored in cupboards or carried around and are only hooked up to the electrical system when they are in use or when they need to be charged up.

A person's use of particular devices will thus be highly varied and dependent on numerous factors ranging from the type of activities they do, i.e. do they cook, do

[1] Most refrigerators offer limited thermostat settings, which are either left at their factory setting or adjusted once in a blue moon.

[2] Some devices offering a standby position draw a small amount of electricity on a constant basis in order to maintain the standby position.

they do laundry, etc., the frequency in which the use their cell phone and thus deplete the battery, whether they live alone or with other people, etc.

Lynch's [15] study of urban environments has shown that people develop detailed mental models of the urban landscape in which they live and that they conceptualize the environment based on the paths they traverse and the landmarks they pass or inhabit during the course of the day.

Do people develop a clear mental model of the electricity infrastructure in their home? Which of the devices with which they interact with stand out as the landmarks within these mental models? Is the electrical infrastructure as invisible as we tend to think or does it factor in people's conception of electricity in their homes?

RQ3 – Which devices people think consume the most electricity?
Not all devices are created equal, when it comes to consuming electricity, some are big in size but miserly, while others are small and nimble yet ravenous. Various factors come into play in determining the amount of electricity consumed by a particular device, the type of operation performed by the device, the length of use and the age are but of few of the factors involved. A proper estimation of the electricity consumed by a given device requires a clear understanding of the factor involved, an understanding that many residential householders may not have.

In their study of how Parisians conceive of their city Milgram and Jodolet [20] found that when people lack personal experience of a given part of a city they will readily use knowledge gained elsewhere or gleaned from the media to create a substitute image that is based as much on fiction as it is on fact.

Studies of people using the London underground system [29] have shown that use of the underground and in particular the development of a mental model of the underground based on a diagram of the system has significantly influenced the way in which they conceptualize the physical structure of the city. At times this misconception led to a huge over or under estimation of distances, severely impacting their navigation and decision-making.

Could the same factors be at play in the home environment with regards to the consumption of electricity? Is the understanding of people based on fact or is it fuelled by misconception?

3 Methodology

To conduct the study we opted for a phenomenological approach inspired by the work of Lynch [15]. Lynch wanted to uncover people's perceptions of the urban environments in which they were living. As part of the interviews he conducted with people he asked them to draw him an "image of the city". The resultant images often naïve portrayals of specific elements and segments of the urban landscape reflected the daily lives of particular individuals. However when taken as a whole and compared to one another the images showed certain commonalities highlighting certain landmarks and marking a common set of boundaries.

As of late a similar methodology was used by Poole, et al [24] in the context of home computer networks, by Yan [30] in understanding how children conceptualize the Internet and by Friedman et al [10] in understanding web security.

As this is our first foray into this domain we decided to first launch a pilot investigation as a means of exploring the lay of the land and establishing the parameters on which further explorations will be based.

The study consisted of a short interview, intended to last for 15-45 minutes in which respondents were asked to draw 3 drawings and after the completion of each drawing describe the objects drawn and the reasons for drawing these objects. Participants were assured that no drawing skills are required and moderators were asked to keep the drawing time to less than a minute per drawing to minimize the potential for contemplation as much as possible as we wanted people to draw what first came to their mind.

At the end of the interview respondents were asked to provide basic demographic information such as age, gender, profession, however no identifying or contact information was collected and thus no follow up interviews were intended nor conducted with the participants.

The three drawings participants were asked to draw corresponded to the research questions posed in the study. For the first drawing respondents were posed with the question “draw me electricity” in an attempt to solicit the first image that came to their mind when they thought about electricity and thus their basic conception of electricity.

For the second drawing respondents were asked to draw all of the electrical elements in their home and for the third drawing they were asked to draw a pictorial representation of the 5 devices they thought consumed the highest amount of electricity.

No video or audio recordings were conducted during the interviews. Following the interviews the objects drawn in each of the drawings were coded and tabulated and summaries created of the types of objects appearing in the drawings.

4 The Drawings

Due to the exploratory nature of the investigation we did not seek to gather a statistically representative sample of the population nor did we concentrate on any particular segment of the population. We simply strove to attract a broad selection of participants from as wide a range of ages as possible in order to provide us with an initial glimpse into the subject domain.

Over a period of 4 weeks a group of 9 undergraduate students who were given training in the method of inquiry canvassed people in and around the city of Funchal, on the island of Madeira, for a brief interview about their perceptions of electricity. Participants were told we were conducting a study on sustainability and were not offered or given any remuneration for their participation.

In all the team interviewed a total of 454 respondents between the ages of 8 to 77 (Mean = 26.55, SD=10.89), of which 50.4% (n=229) were male and 49.5% (n=225) were female.

Due to the predominance of young individuals, most of whom were students or young professionals in the sample, no attempt was made to analyze the results with regards to age or profession. The following is a preliminary analysis of the results.

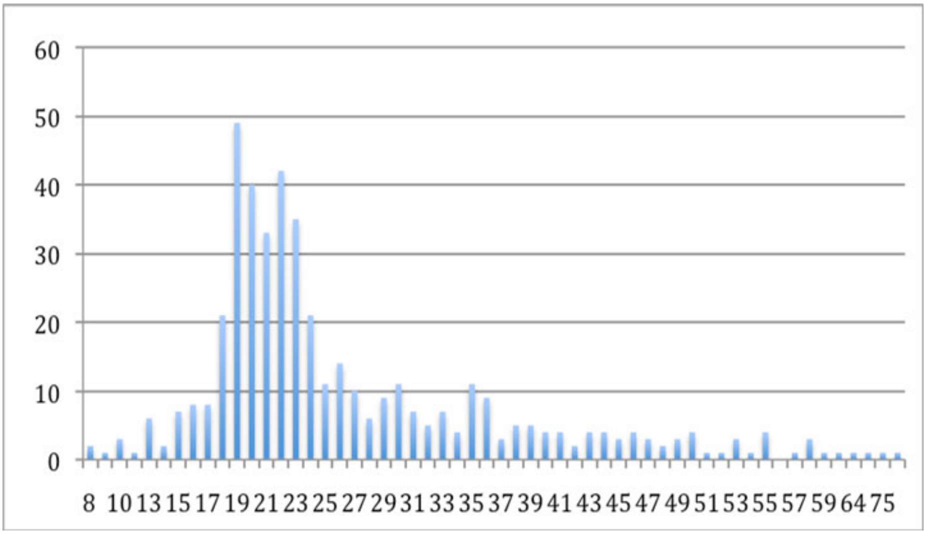


Fig. 1. Age breakdown of the respondents

4.1 Drawing 1 – Draw Me Electricity

In the first drawing participants were asked to draw the first thing that came to their mind when they thought of electricity. However in some cases the participants got carried away and drew several elements and the moderators caught in the heat of the moment failed to mark which of the elements was the first one to be drawn.

In order to maintain consistency when comparing the drawings, drawings containing more than one element were removed from the sample leaving a total of 400 drawings. The age and gender distribution within this sub-set remained largely unchanged. The age of the respondents ranged from 8 to 77 with a mean age of 26.19 (SD=10.76) of which 50.5% (n=202) were females and 49.5% (n=198) males.

The most commonly drawn object was a light bulb with 39.5% (n=158) of all drawings. An additional 6% (n=24) of participants drew various light fixtures making light the most common representation of electricity with 45.5% (n=182) of all drawings. There was a marked difference between the genders here with 63.7% (n=116) of the drawings drawn by females and 36.2% (n=66) being drawn by males.

When asked why they drew a light bulb or a light fixture all of the participants irrespective of gender remarked that electric lights are the most prevalent manifestation of electricity. Statements like “we are surrounded by light”, “electric lights are everywhere”, “light is the first thing I see” and “turning on the lights is the first thing I do” were repeated over and over again.

The second most common object drawn was Lightning with 19.25% (n=77) of the drawings. The majority drew simple representations of lightning bolts although a few drew more elaborate affairs portraying a thunderstorm complete with clouds and rain. When queried about the reasoning behind the drawing a large number said that although their initial thought was of a thunderstorm, the lightning bolt had a double meaning as it is often used as a warning sign on electrical installations. This set of drawings also displayed a marked difference between the sexes but in a reverse order with 77% (n=59) being drawn by males and 23% (n=18) by females.

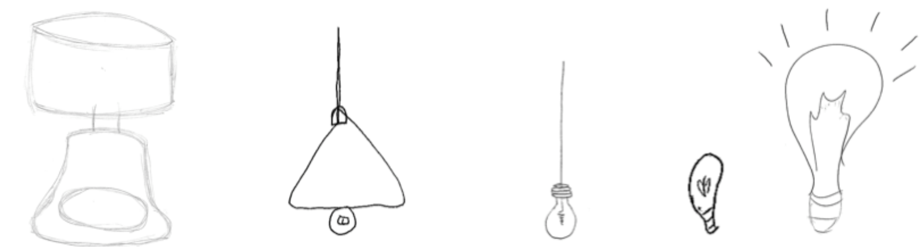


Fig. 2. A representative sample of the light bulbs and light fixtures drawn by the respondents

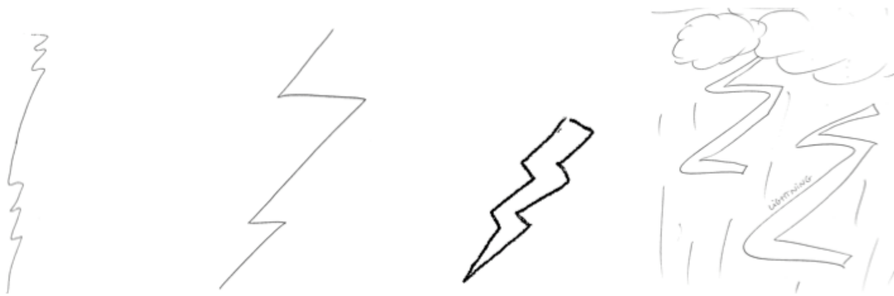


Fig. 3. A representative sample of the lightning bolts drawn by the respondents

Fig. 4. A representative sample of abstract and diagrammatic representations of an electric current drawn by the respondents

In third place with 8.5% (n=34) were drawings of an electrical current drawn in either a purely abstract from (n=20) or in a diagrammatic format showing an electrical circuit (n=14). Of these 70.6% (n=24) were drawn by males and 29.4% (n=10) were drawn by females. The drawings created by the male respondents were equally divided between the abstract and diagrammatic formats while those drawn by female respondents were mainly abstract (n=8)

Electrical infrastructure elements such as electricity poles, cables and sockets were next in line with 7.5% (n=30) of participants divided equally between males and females drawing an electricity pole and 5% (n=20) drawing indoor power cables and plugs. Only 2% (n=4) of participants drew a wall socket and 0.75% (n=3) drew a light switch.

Fig. 5. A representative sample of the electricity infrastructure elements drawn by the respondents

The remaining drawings (n=33) representing 8.25% of the total reflected a variety of notions from the concrete to the abstract. Two respondents drew a person with spiked hair and another drew a person being shocked while touching a surface as a representation of static electricity. Three respondents drew the symbol of the Euro as a representation of the cost of electricity. Two additional respondents drew batteries while another pair drew a representation of the flow of electrons in a wire as a representation of the essence of electricity. Moving towards the purely abstract one person drew a butterfly because it represents energy and colour and another drew a nightscape as "it is in the night when electricity is most needed"

Fig. 6. Some of the more abstract representations drawn by the respondents, pictured from left to right, the flow of electrons thru a wire, a pictorial representation of the night, a person with statically charged spiked hair, and a butterfly drawn to represent energy and colour.

It is interesting to note that not a single respondent drew an image of a power station (with or without smoke stacks) although 3 of the respondents drew windmills and indicated their use as sustainable energy source, 3 others drew an image of the sun and indicated its potential as an alternative energy source and one respondent drew a representation of a hydro electric dam.

4.2 Drawing 2 – Electricity in the Home

In this drawing respondents were asked to draw electricity in their home. We sought to see the way in which people perceive and conceptualize electricity within their homes and whether the hidden and embedded infrastructure elements such as the electricity wires that snake through the walls as well as the sockets and switches that slightly protrude from the walls would be reflected in these perceptions. We suspected that infrastructure elements would by and large be hidden and that devices people use on a frequent basis would figure prominently in their perceptions.

Figure 7 provides an overview of the type of drawings created by the respondents. The vast majority of respondents (79.42%, n= 359) opted to simply draw various devices and elements on the page with no specific relationship between the objects drawn. A further 11.95% (n= 54) drew floor plans of their homes and positioned the elements in their relative position in the home while another 8.63% (n=39) drew a stylized representation of their home and placed the entities within the home environment. Female respondents were more likely to draw floor plans (n=38) or stylized representations (n=28) than males who only drew 16 floor plans and 11 stylized representations. However the number of elements portrayed in each drawing was roughly the same with males drawing an average of 3.72 (SD=2.46) per drawing and females drawing an average of 3.83 (SD=2.64) elements per drawing.

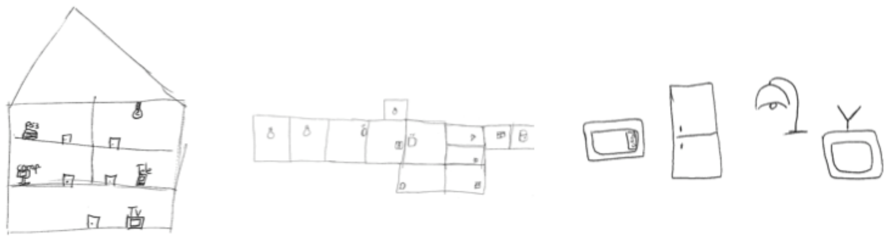


Fig. 7. A sample of the drawings created by the respondents for drawing 2

The drawings (n=452) confirm our suspicion that infrastructure elements are by and large out of sight and out of mind, however they are not completely invisible with 24.11% (n=109) of the drawings containing at least one infrastructure element as illustrated in Table 1. Males appear to be more sensitive than females to infrastructure elements built into the walls such as wall sockets and wires but the their representations of the fuse box and the electricity meter appear to be more analogous.

Table 1. Electricity Infrastructure Elements in Drawing 2

Element	Female	Male	Total	Percentage
Wall socket	28	50	78	17.26
Fuse box	12	14	26	5.75
Wire	4	7	11	2.43
Power strip	4	6	10	2.21
Electricity Meter	3	4	7	1.55

Electric lights figure prominently in the drawings with 46.20% (n=208) containing a representation of a lamp, a light bulb or a light switch. However lights also present a classification problem as some lights such as ceiling lights are built into the walls and are turned on from switches purposely built into the walls and thus may rightly perceived as part of the infrastructure of the house and others such as bed side lamps are stand alone devices which are turned on from a switch on the lamp and thus may be perceived as an independent device. Only 13.27% (n=60) contained a specific representation of a table or a floor lamp and they appeared more frequently in drawings created by females (n=36) than by males (n=24). 15.04% (n=68) of the drawings contained a drawing of a light bulb, with the bulbs appearing more frequently in drawings created by male respondents (n=44) than those created by female respondents (n=24). Ceiling lights drawn in either great detail or in purely abstract form appeared in 19.92% of the drawings (n= 90) with a higher preponderance in the drawings created by females (n=55) to those created by males (n=35). Light switches appeared in 5.31% (n=24) of the drawings with a fairly equal distribution between the genders with 13 drawn by males and 11 by females. In light of these data we are inclined to classify light as an infrastructure element.

The drawings also support our second hypothesis that devices people use on a frequent basis would feature prominently in their perception of electricity. Electric lights are a case in point however televisions appear to shine more brightly in the perceptions of the respondents with 52.21% (n=236) of the drawings containing a representation of a television set and computers are not far behind appearing in 41.81% (n=189) of the drawings. Large domestic appliances loom large in the drawings appearing at least once in 54.65% (n=247) of the drawings. Refrigerators appear most frequently in the drawings followed by Microwave ovens and washing machines. Small appliances such toasters, juicers and irons feature in only 21.46% (n=97) of the drawings, most appear in only a few drawings and none have broken the 10% barrier.

An interesting anomaly but an understandable one given Madeira's mild weather is the almost total absence of heating and cooling devices in the drawings. In all only 1 air conditioner, 3 heaters, 4 fans and 4 dehumidifiers were drawn with the heaters and fans drawn mainly by males and the air conditioners and dehumidifiers mainly by females. A more interesting and less understandable anomaly is the low number of cell phones drawn with only 6.64% (n=30) containing a representation of a cell phone and the almost complete absence of iPods/MP3 players with only 3 iPods drawn despite the high prominence of these devices in everyday life and the high frequency in which they need to be charged. One explanation might be that these devices have become so ubiquitous that they, and the activities needed to support them have receded into the background.

Table 2 provides an overview of the devices that appear most frequently in the drawings created by male and female respondents. The percentages are computed per gender.

Table 2. Electrical Devices Featured Most Frequently in Drawing 2

	Female			**Male**		
	Device	**Count**	**%**	**Device**	**Count**	**%**
1	Television	120	53.81	Television	116	50.66
2	Computer	86	38.57	Computer	103	44.98
3	Refrigerator	85	38.12	Refrigerator	77	33.62
4	Microwave	62	27.80	Microwave	55	24.02
5	Washing machine	41	18.39	Washing machine	35	15.28
6	Toaster	21	9.42	Radio	17	7.42
7	Hair dryer	20	8.97	Cell phone	14	6.11
8	Radio	16	7.17	Iron	14	6.11
9	Cell phone	16	7.17	Oven	13	5.68
10	Iron	15	6.73	Stereo system	11	4.80

4.3 Drawing 3 – Draw the Highest Consuming Devices

In this drawing respondents were asked to draw representations of 5 electrical devices in their homes they thought consumed the highest amount of electricity.

Although an extensive range of appliances and devices was drawn ranging from Aquariums, air compressors, alarm clocks, curling irons and cement mixers through electric guitars, humidifiers and foot massagers to sewing machines, soldering irons and water coolers those that appeared most frequently were of a more mundane nature as listed in table 3. Unfortunately the order in which the devices were drawn was not recorded so we cannot say which device was drawn first, second and so forth and only report on the total occurrence of specific devices in the drawings.

The drawings and the comments made by the respondents reflect the notion that people do not have an accurate sense of the amount of electricity consumed by a particular device and tend to associate consumption level with the frequency and duration of use as well as with the size of the device.

Table 3. Top 10 Most Frequently Drawn Devices in Drawing 3

Female			Male		
Device	**Count**	**Percentage**	**Device**	**Count**	**Percentage**
Television	141	62.67	Refrigerator	154	67.25
Refrigerator	139	61.78	Television	145	63.32
Washing machine	131	58.22	Computer	127	55.46
Computer	112	49.78	Washing machine	121	52.84
Microwave	92	40.89	Microwave	102	44.54
Iron	61	27.11	Iron	56	24.45
Deep freezer	57	25.33	Light	55	24.02
Light	48	21.33	Deep freezer	42	18.34
Water heater	34	15.11	Oven	32	13.97
Tumble Dryer	32	14.22	Water heater	28	12.23

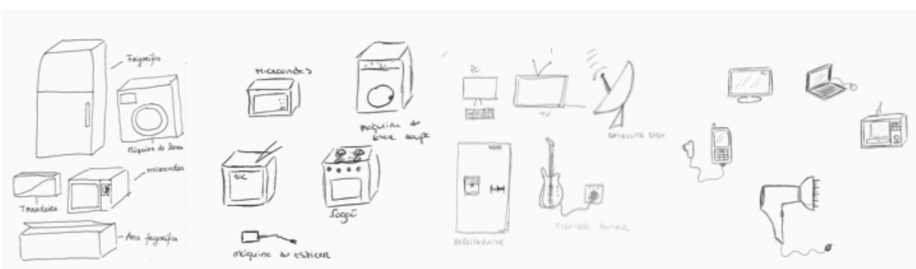


Fig. 8. A Sample of the drawings created by the respondents for drawing 3

Most of the devices listed in Table 3 consume low to moderate levels of electricity but when queried about why they drew a particular device respondents constantly made references to size, duration and use, with statements like, “the television is on all the time so it must be consuming lots of electricity”, “I have lots of lights in my house”, “we have a big deep freezer”, “the kids use the computer all the time”.

In one extreme example a 23-year-old male drew a small aquarium, a fridge, a deep freezer, a television and a computer. When queried about the drawing he said the water pump and the light in the aquarium are on all the time, he watches a lot of television and spends a lot of time with the computer and the refrigerator and deep freezer are very big and for these reasons they each consume a lot of electricity.

A few respondents drew a person in addition to devices to highlight the high rate of consumption brought about by a particular individual. An example can be seen in the drawings of two respondents. The first, a 28-year-old male who in addition to a computer, refrigerator, iron and a television, drew his girlfriend, and noted that she never turns off the lights in his apartment. The second, a 25-year-old female, drew her grandmother in between a television (intended to represent the 4 TV sets in her house), a computer, a cell phone (intended as a representation of the 5 cell phones in her house) and a microwave and complained that the old lady uses the washing machine and tumble dryer on a daily basis and thus consumes a lot of electricity.

Fig. 9. A tale of an aquarium, a girlfriend and a grandmother

4.4 Some Initial Conclusions

Although the results of the study were drawn from the population of the city of Funchal and its outlying areas and therefore represent the perceptions of an urban population with a predominantly European background we believe that they provide a valuable insight into people’s perceptions of electricity on which further study and reflection can be based.

The drawings confirm our suspicion that people lack a coherent mental model of electricity. The hidden and mediated nature of electricity means that people’s perceptions of electricity are shaped by the devices that pervade their daily lives, and, as a result electric lights, computers, televisions and a host of domestic appliances dominate individual perceptions of electricity. Without visual cues to the amount of electricity consumed by each device people seem to rely on mental models and associations gleaned elsewhere. Thus, the size of the device, the duration of use and the frequency of use become proxies for rates of consumption. As a result people ascribe high consumption levels to the most miserly of devices, to the extent that even the lowly aquarium light may be perceived as an electricity guzzler simply because it is on all of the time.

An examination of the devices listed in Table 3 reveals only minor difference in the perceptions of males and females with with regards to which devices consume the highest amount of electricity. Thus although the males of the species may not do as many loads of laundry as the females, they both attribute the device with a high level of consumption. After all frequency and size are in the eyes of the beholder, regardless of whether the beholder actually uses the machine or beholds an image of someone else using the machine as illustrated in Figure 10.

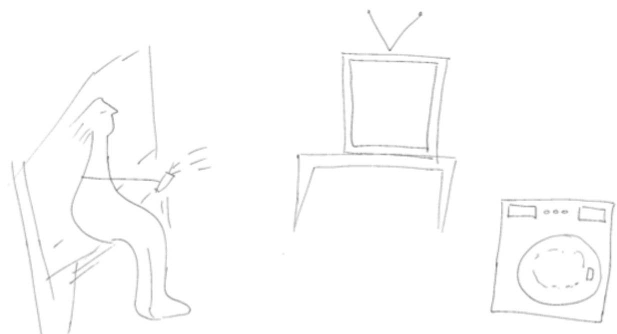


Fig. 10. Still life with husband, television and washing machine. A drawing of electricity in the home created by a 36 year old, female

5 Discussion

In *Understanding Media*, McLuhan [18] famously quipped, "The medium is the message". In the same volume he used electric light as an example stating, "The electric light is pure information. It is a medium without a message…". McLuhan wanted to stress a point, that the impact (message) of a technology (media) was not what the technology enabled, or even what one did with the technology but what were the social ramifications of the technology and the patterns of use that followed its adoption.

To paraphrase McLuhan's example the importance of the railway is not in the way it harnessed rails, wheels and steam power to move people and cargo, or in the specific trips individual people made but in the new types of urban forms and new forms of work and leisure activities that have emerged.

In this sense the light bulb is indeed a medium without a message unless the message is one of convenience and abundance. However we must ask ourselves whether the social ramifications of a technology convey the whole message or are they just one aspect of the story.

A light bulb emits light, that is plain to see, but if we hold our hand next to it, we can sense that it also emits heat. The heat we can feel being emitted from the lamp and the emissions released by the power station that powers it are costs whose ramifications have largely been ignored until now.

Most current sustainability related efforts have been geared towards bringing electricity and the hidden costs associated with its consumption to the attention of consumers in an effort to elicit a more sustainable lifestyle. However as the drawings reveal we are not only facing a problem of visibility we are also facing a problem of legibility. By legibility we essentially mean the ease in which people interpret and understand a place, a technology or a piece of information. The challenge now is how to make the consumption of electricity not only more visible but also more legible. The issue of legibility raises a number of interesting questions.

1. How can we imbue electricity and its consumption with a meaning people can easily perceive and relate to?
2. How can we relate the consumption rate of individual devices in a way that would enable people to develop a mental model based on the properties of a device, i.e. does it provide light, heat, refrigeration, etc. as opposed to its size, duration or frequency of use?
3. How can we utilize the notions of size, frequency and duration in promoting more sustainable use and consumption patterns?

As noted above one of the difficulties with the consumption of electricity is its invisible nature. People tend to have difficulty conceptualizing invisible or purely abstract concepts such as electricity, preferring instead to rely on visible manifestations and physical properties such as the light emitted by a bulb or the size of a refrigerator.

One answer to the above questions would be to use light and the size of a device as elements in design. Gustafsson & Gyllenswärd [12] successfully made manifest the flow of electricity through a cord by making the cord glow with light. Why not extend the metaphor and make the whole refrigerator glow with light. By contrasting the large size of the refrigerator with a low level of ambient light (since refrigerators are relatively low consumers of electricity) we will hopefully enable users of the refrigerator to make the correct relationship between size and consumption level. A whole house equipped with such appliances will enable the residents to map out their electrical consumption and focus their conservation efforts on particular problem spots.

Further playing with the notions of size, duration and frequency of use which seem to dominate the perceptions of the study participants we can use Non Invasive Load Monitoring (NILM) devices to get a detailed record of the electrical consumption of various devices in the house and then display graphical representations of the consumption levels of these devices coupled with pithy quirky messages. For example, a television may display the following message upon stating up “I maybe a 42” television screen but I only consume as much electricity as the toaster you used this morning”.

Fig. 11. The consumption and production of electricity in the home as portrayed by a 30 year old female respondent. Electricity is consumed by a host of domestic appliances; the consumption is then offset by electricity generated as a result of amorous activities in the marital bed.

6 Conclusion

In *The Image of The City*, the work that provided the initial inspiration for this study, Lynch [15] strove to discover how people understood the structure of cities, what were their perceptions of specific areas of the city, how did they navigate through the space and how would they define and portray its properties to others. Lynch's objective was to understand the *legibility* and *imageability* of urban places, properties that would not only allow the inhabitants to easily navigate through the city but that would also foster a local identity and pride of place and that would allow urban planners and developers to create urban environments that are liveable as well as habitable.

In this work we strove to follow in Lynch's footsteps by trying to uncover people's perceptions of electricity in order to devise methods that would make electricity more legible and imageable (to use Lynch's terms) and thus support a more sustainable use of electricity, one that is based on understanding of the benefits and consequences of the consumption of energy as opposed to one based on indifference, ignorance and guilt.

Although the study was conducted within the confines of a small island with a fairly homogenous population we believe the work provides valuable insights into how people perceive electricity and how perceive the consumption levels of individual devices and thus to be of use both to HCI researchers looking to develop sustainable technologies and technologies intended to support sustainability and researchers in other fields engaged in projects related to electricity and sustainability.

We intend to pursue this work further by conducting additional studies in which we will seek both to validate the findings by surveying a more diverse population base on which we could conduct rigorous statistics test and to expand the scope of the work by looking at additional aspects such as the electricity grid and the energy sources used to power it.

We would like to end this paper with a final image. An image we believe portrays both the rich insights on electricity provided by this study, insights we have only begun to explore and thus only cursorily presented in this paper and the power of drawing as a tool for eliciting rich information on recalcitrant topics from reserved as well as responsive respondents.

Acknowledgements. Support for this research was provided by the Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology) through the Carnegie Mellon Portugal Program as part of SINAIS – Sustainable Interaction with social Networks, context Awareness and Innovative Services (CMU-PT/HuMach/0004/2008). Additional funding was provided by CITMA - Centro de Ciência e Tecnologia da Madeira (The Madeira Center for Science and Technology). We would also like to thank the students who helped conduct the fieldwork for this study; Angela Ferreira, Catarina Dias, Elvio Pita, Francisco Andrade, Francisco Capelo, Maurilia Nobrega, Nicolene Baeta, Yesica Reynolds, and Wilmer Goncalves; all of the participants who contributed their time, patience and drawing skills to this effort; and the anonymous reviewers who provided helpful feedback on an earlier draft of this paper.

References

1. Attari, S.Z., DeKay, M.L., Davidson, C.I., de Bruin, W.B.: Public perceptions of energy consumption and saving. In: Proceedings of the National Academy of Sciences of the United States of America (2010); PNAS published ahead of print August 16, 2010, Retrieved from:
`http://www.pnas.org/content/early/2010/08/06/1001509107.full.pdf+html`
2. Bostrom, A., Morgan, M.G., Fischhoff, B., Read, D.: What do people know about global climate change? 1. Mental models. Risk Analysis 14(6), 959–970 (1994)
3. Chetty, M., Banks, R., Harper, R., Regan, T., Sellen, A., Gkantsidis, C., Karagiannis, T., et al.: Who's Hogging The Bandwidth?: The Consequences Of Revealing The Invisible In The Home. Technology, 659–668 (2010)
4. Chetty, M., Brush, A.J.B., Meyers, B.R., Johns, P.: It's not easy being green: understanding home computer power management. In: Proceedings of the 27th International Conference on Human Factors in Computing Systems (CHI 2009), pp. 1033–1042 (2009)
5. Chetty, M., Tran, D., Grinter, R.E.: Getting to green: understanding resource consumption in the home. In: Proceedings of the 10th International Conference on Ubiquitous Computing (UbiComp 2008), pp. 242–251 (2008)
6. Darby, S.: The Effectiveness of Feedback on Energy Consumption. A Review for DEFRA of the Literature on Metering Billing and direct Displays. Oxford University – Environmental Change Institute (2006), Retrieved from:
`http://www.vaasaett.com:81/xmlui/handle/123456789/133`
7. Dillahunt, T., Mankoff, J., Paulos, E., Fussell, S.: It's not all about "Green": energy use in low-income communities. In: Proceedings of the 11th International Conference on Ubiquitous Computing (Ubicomp 2009), pp. 255–264 (2009)
8. Dunlap, R.E.: Lay Perceptions of Global Risk: Public Views of Global Warming in Cross-National Context. International Sociology 13, 473–498 (1998)

9. Foster, D., Lawson, S., Blythe, M., Cairns, P.: Wattsup?: Motivating reductions in domestic energy consumption using social networks. In: Procedings of Nordi CHI 2010 (2010)
10. Friedman, B., Hurley, D., Howe, D.C., Felten, E., Nissenbaum, H.: Users' Conceptions of Web Security: A Comparative Study. In: CHI 2002 Extended Abstracts. ACM Press, New York (2002)
11. Froehlich, J., Findlater, L., Landay, J.: The design of eco-feedback technology. In: Proceedings of the 28th International Conference on Human Factors in Computing Systems, CHI 2010 (1999)
12. Gustafsson, A., Gyllenswärd, M.: The power- aware cord: energy awareness through ambient information display. In: Ext. Abs. CHI 2005, pp. 1423–1426. ACM Press, New York (2005)
13. He, H.A., Greenberg, S.: Motivating Sustainable Energy Consumption in the Home. Psychologist, 1–5 (2008)
14. Furby, L., Slovic, P., Fischhoff, B., Gregory, R.: Public Perceptions of Electric Power Transmission Lines. Journal of Environmental Psychology 8, 19–43 (1988)
15. Lynch, K.: The Image of the City. MIT Press, Cambridge (1960)
16. Mankoff, J., Matthews, D., Fussell, S.R., Johnson, M.: Leveraging Social Networks to Motivate Individuals to Reduce their Ecological Footprints. In: Proc. HICSS 2007 (2007)
17. Martin, N.: Habit: The 95% of Behavior Marketers Ignore. FT Press, Upper Saddle River (2008)
18. McLuhan, M.: Understanding Media: The Extensions of Man. MIT Press, Cambridge (1994)
19. Mehlwana, A.M.: The anthropology of fuels: situational analysis and energy use in urban low-income townships of South Africa. Energy for Sustainable Development 3(5), 5–15 (1997); International Energy Initiative, Inc., Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S0973082608602082
20. Milgram, S., Jodelet, D.: Psychological maps of Paris. In: Proshan, H., Ittelson, W., Rivlin, L. (eds.) Environmental Psychology, 2nd edn., pp. 104–124. Holt, Rinehart & Winston, New York (1976)
21. Pierce, J., Odom, W., Blevis, E.: Energy aware dwelling: a critical survey of interaction design for eco-visualizations. In: Proceedings of the 20th Australasian Conference on Computer Human Interaction - Designing for Habitus and Habitat, pp. 1–8 (2008)
22. Pierce, J., Paulos, E.: Materializing energy in everyday life. In: Proceedings of CHI 2010 (2010)
23. Pierce, J., Schiano, D.J., Paulos, E.: Home, Habits, and Energy: Examining Domestic Interactions and Energy Consumption. Human Factors, 1985–1994 (2010)
24. Poole, E.S., Chetty, M., Grinter, R.E., Edwards, W.K.: More than meets the eye: transforming the user experience of home network management. Designing Interactive Systems, 455–464 (2008)
25. Read, D., Bostrom, A., Morgan, M.G., Fischhoff, B., Smuts, T.: What do people know about global climate change? 2: Survey Studies of Educated Laypeople. Risk Analysis 14(6), 971–982 (1994)
26. Rittel, H., Webber, M.: Dilemmas in a General Theory of Planning. Policy Sciences 4, 155–169 (1973)
27. Sengers, P., Boehner, K., David, S., Kaye, J.: Reflective design. In: Proc. of Critical Computing 2005, pp. 49–58. ACM Press, New York (2005)
28. Strengers, Y.: Smart metering demand management programs: challenging the comfort and cleanliness habitus of households. In: Proceedings of the 20th Australasian Conference on Computer Human Interaction, OZCHI 2008, pp. 9–16 (2008)
29. Vertesi, J.: Mind the Gap: The "Tube Map" as London's User Interface. In: Engaging the City Workshop Paper, CHI 2005 (2005)
30. Yan, Z.: What Influences Children' s and Adolescents' Understanding of the Complexity of the Internet? Developmental Psychology 42(3), 418–428 (2006)

Structuring the Collaboration of Multiple Novice Design Ethnographers: Towards a New User Research Approach

Paul Gault[1], Catriona Macaulay[1], Graham Johnson[2], and Judith Masthoff[3]

[1] School of Computing,
University of Dundee,
Dundee, Angus, DD1 4HN, UK
{p.e.gault,c.macaulay}@dundee.ac.uk
[2] Advanced Technology and Research, NCR Labs,
NCR Corp, Dundee, DD2 4SW
graham.johnson@ncr.com
[3] University of Aberdeen,
Meston Building,
King's College, Aberdeen, AB24 3UE
j.masthoff@abdn.ac.uk

Abstract. This paper proposes a new design research method to support businesses engaging in the innovation of products and services intended for use in public spaces. Increasing numbers of companies are turning to detailed user/consumer research often based on ideas from the design ethnography community. In an increasingly complex and fast moving business world, there is a need for faster user research that also provides a wider focus on the situation under investigation. A potential solution is using a larger number of fieldworkers on one study. As it would be difficult and costly to utilise many experienced design ethnographers, this may also involve the use of novices. This paper describes the development of a method for adapting existing practices to the emerging context outlined above (i.e. large numbers of fieldworkers, not all of whom necessarily have experience in ethnography). We discuss 3 field studies that show how the method can be applied and how it has been fine-tuned based on the outcomes. This method involves multiple groups of fieldworkers situated at a range of public spaces and each assigned with a specific theme of interest. The wealth of material that this fieldwork activity produces is then digested and insights are generated from it to help inform an understanding of existing behaviour within public space. This paper shows that fieldwork can be reduced to a set of simple tasks that can be successfully distributed over a group of novices facilitated by an experienced design ethnographer. This work will be extended further so that it can be applied as part of a toolkit for use in businesses where there is no established culture of utilising this type of user research.

Keywords: Ethnography, Crowdsourcing, Fieldwork, Innovation, Methodology, Public Space.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 118–135, 2011.

1 Introduction

We will address the problem of providing ethnographic research for an increasingly complex and fast moving business world. This type of activity usually takes a long time to perform and is very specific in what it is observing. Our aim is to create a methodology to help explore the problem space in a short space of time with people who do not necessarily have a background in design ethnography. An objective of this is to have a broad focus on the subject matter so that there is a wide range of perspectives on what is being observed.

As a case study, we will try to identify opportunities for collaborative technologies situated within public space that encourage social interaction. Some examples of the kinds of new systems researchers and companies are interested in are digital signage, ubiquitous computing and mobile technologies. As public space is such a large context, we require a solution that will help us to understand the huge variety of activity going on there at any one time. Instead of a small number of fieldworkers taking many weeks to understand each of the perspectives of people who inhabit the space, a potential solution is to gain multiple viewpoints from many people making observations at the same time. We created a set of very broad themes that could be interpreted in multiple ways so that there would be a variety of approaches to gathering the fieldwork data. We will discuss the design and evaluation of the materials used to facilitate the fieldwork activity.

In essence, we will suggest a method for distributing fieldwork across a group of novice observers, reducing the fieldwork activity to a set of simple tasks. These tasks will be framed and facilitated by an experienced design ethnographer who will produce the support materials and lead the analysis workshops. The idea of using a group of people in fieldwork has similarities with Open Source development [1, 2]. In Open Source development, the community is encouraged to apply their skills to extend existing software. Our proposed methodology will provide more structure to the tasks requiring less expertise from participants. The idea of involving non-experts in fieldwork has similarities with crowdsourcing [3]. In crowdsourcing, well-defined tasks that are traditionally performed by employees are outsourced to an undefined large group of people via an open call. Our approach will not necessarily use an open call to recruit participants, and will involve more explicit continuous collaboration between the participants.

To illustrate our method, we will discuss a set of studies using observations of the interactions taking place in social spaces such as shopping centres, city squares and railway concourses. The outcomes of these studies provide an insight into how to design for such contexts along with techniques for managing this process.

In section 2 we will describe the other relevant literature that has helped to inform the development of our methodology. In section 3 we provide an overview for each of the main activities that constitute up our method. We explain how the method was first deployed in section 4 then subsequently refined and applied an additional two times in section 5. Finally, in section 6 we will detail the conclusions that were arrived at and indicate potential future work.

2 Related Work

Design ethnography requires a set of custom techniques that can be adapted from their roots in anthropology to suit the requirements of new product development [4]. Previous studies have shown the application of ethnographic techniques using short timeframes [5] but the possibility of working with a number of fieldworker observations has not been explored.

We aim to use fieldwork to drive a broad exploration of the problem space. The analysis and synthesis of fieldwork observations will help to uncover new opportunities for the interactive systems we are interested in.

The normal limitations of fieldwork-based techniques include issues related to time, breadth and generalisability [6]. Fieldwork can be a very time consuming approach; doing both the observations and analysing the gathered data. It lacks breadth as it typically only studies one organization or culture in a specific time period, so offering in-depth knowledge only of particular contexts and situations. The generalisations arrived at from one study may be subjective as the researcher may have imposed an interpretation on what has been observed (different people may observe different things). We propose to use this subjectivity to our advantage as a way of channelling the observations from a variety of perspectives.

The initial stage of the methodology is the gathering of the fieldwork data. This will involve sending our participants out into a large public space with a set of themes that can help to guide the observations they make. Similar approaches [7] have shown that this technique will support the fieldworkers as they gather material by focussing on a specific topic of interest. Each theme will be explained using a reference card alongside a map of the physical space that they must make their observations within. The usage of reference cards to help explain a design methodology is a well-established concept [8]. We will use this technique as a way of co-ordinating the observations made by our group of novice fieldworkers.

For the second part of our methodology, we will undertake an affinity-mapping workshop where we analyze the fieldwork data using the initial steps of a grounded theory approach, namely open coding [9]. This is appropriate, as it will allow new themes to emerge from each fieldworker's individual field notes. The process of combining these themes will also allow us to arrive at a set of new-shared themes derived from all of the participants' fieldwork data.

Bringing together all of the field notes from the individual fieldworkers will involve the creation of an affinity-mapping diagram. Affinity mapping is originally based on the KJ-method [10], which develops themes into insights and helps identify the relationships among many issues. This technique will enable the group to consider the results of all of the observations made during the fieldwork activity. The outcome of this will be a KJ image that shows an externalized mental map of the fieldwork observations made by the group. This will help to show the broad themes relevant to the problem alongside the data they are derived from. This is a useful process as, by its nature, it is an activity that must be done fast encouraging a systemic response to the overall problem [11] instead of slowly going through with point-by-point solutions to each problem that is identified. During the workshops we will use the physical space to support the collaborative analysis work. The ability for large media to support many people collaborating in the same physical space [12] will help our participants arrive at a shared understanding of the fieldwork material they have collected. This will enable them to discuss the detail of each individual's fieldwork notes and relate this to the material that other participants have contributed.

In our method, we ask broad exploratory questions first. When working in an industry setting, it can be very difficult to be very specific at the outset. For example, "find the next killer application" might be the first question that a company asks when exploring a new area. By keeping it as broad as possible initially, this gives us enough space for any potential solutions (including unexpected ones) to be explored. The data analysis may result in more specific questions that can then be investigated in a new round of fieldwork.

3 Proposed Methodology

We will provide a brief summary of the method and then we will go into greater detail by explaining each of the main stages that were followed. Firstly, we broke down our problem of identifying opportunities for collaborative technologies situated within public space into 4 themes. We prepared materials for the participants to refer to whilst out in the field and assigned 2 person teams to work on each theme and placed them at different locations. The fieldwork observations took place on a single day where a large amount of data was recorded. This material was then brought back into the studio, shared with other group members and consolidated into a secure online space. Finally, a workshop was held where the group ruminated over the material they had gathered and some directions for further research exploration were identified. The four main phases of activity were:

- Materials preparation
- Fieldwork observations
- Dump & distil
- Affinity mapping workshop

3.1 Materials Preparation

A set of briefing cards was designed for the participants to refer to whilst out in the field making their observations. This gave each of the teams a key area to focus on both conceptually and physically. One side of the card detailed a research theme with a title and a set of questions whilst the reverse of the card had a map marking out where the observations were to be located. The research theme was a title followed by single sentence along with 3 example questions loosely related to the theme. The cards were A5 size so they were small enough for the participants to keep in their pocket and so easily referred to if needed. The cards were also laminated to make them robust in case the weather conditions were not favourable.

3.2 Fieldwork Observations

This phase of work is where the participants go out into the field and make their observations. The participant group were split into equal numbers of people and each team was given a briefing card containing a specific theme. The whole group were briefed before setting off to make their observations. It was suggested that one person take the role of writing notes and another person of capturing visual material. If it was felt necessary, the team members were free to switch roles at any point during the exercise. The importance of keeping communication flowing between members was stressed so that there was a shared plan of action for what was being observed. The teams were encouraged to take plenty of breaks, as making the observations can be quite an intensive exercise. During these breaks, they were able to use the opportunity to talk about how the task was progressing. The teams were asked to keep any observations in context so that they were mindful of the environmental conditions, time of day and density of people in the space they were observing.

3.3 Dump and Distil

This is the process of putting all the fieldwork data into a form so that it can be worked with further. It began as soon as the observation activity was completed. The teams were asked to write-up another version of the scratch notes they had recorded immediately after making their observations so that nothing was forgotten, as this would easily happen otherwise. The whole group came together to discuss their observations and reflect upon the process they had gone through whilst out in the field. From the outset, the group were encouraged to remain analytical about their observations by not jumping ahead into thinking about designing new solutions. They were asked to discuss their strategy throughout the task, how they interpreted the individual research questions and how they approached working in each location. The group were guided to create backups of all their notes in case they were lost and also, they could be easily shared with each other. To facilitate this, secure network drives were created where everyone was able to pool their data for further analysis.

3.4 Affinity Mapping Workshop

The final part of our method involved a workshop to help reach a shared understanding of the material that had been gathered. The teams were asked to make a pass through their fieldwork data and extract items of interest onto post-it notes. Each team was guided to use 1 colour of post-its and initial each note so that if necessary, the data could be queried later on in the process. Once the groups had written out their notes, they started to arrange them into sub-groups and used index cards to write headings above the groupings. A photograph of each team's notes was taken at this point so if needed, the layout of notes was saved and would be referred back to later on. The final part of the workshop involved all of the teams pooling their data together to see what overall findings could be arrived at. This involved taking down the sub-headings from each group and arranging these into super-headings across all of the groups. The original post-it notes were pulled into these new super-headings so it could be seen how appropriate they were. It should be stressed that this activity can be repeated a number of times to see if the same or a different understanding is reached.

4 Field Study 1

4.1 Study Design

Following the results of our previous Urban Traces study [13], we decided to develop our methodology by creating a number of themes and targeting specific physical spaces related to them.

4.1.1 Questions. The research questions for the fieldwork observations were split into 4 different themes. Each theme was explained using a sentence outlining the focus of the observations and then 3 follow-up questions that the participants were asked to try to answer. The themes, explanations and questions were as follows:

Connections. Observe how people navigate their way through physical public space using the way finding systems that are present. How useful are the existing way finding systems? Where do clusters of people and flows of activity occur? What would happen if the way finding system were to fail?

Transactions. Observe how the exchange of information, goods & services is facilitated. What are the social norms and practices taking place around self-service kiosks? How are people using public space as a platform for collaboration? Where is technology being used and why?

Interactions. Observe how people interact both with each other and with the systems that are present in public space. How does the space affect the way in which people behave? Where is one-to-one & one-to-many communication occurring? What kinds of technology are present and why are they there?

Conversations. Observe the patterns in behaviour for ways in which people congregate in public space. What are the differences between how large groups of people behave in comparison to small groups & individuals? How does the physical layout affect places that people gather? Where are the empty spaces and why are they so?

4.1.2 Participants. Fourteen students on a Masters of Design Ethnography (MDE) participated in the study (4 male, 10 female). The course teaches students to perform user research and explore how this is incorporated into the work of companies globally. The participants were grouped into teams of 2 so that the different activities could be shared out within each team.

4.1.3 Location. The city centre of Glasgow was used to carry out the fieldwork activity. Glasgow is the largest city in Scotland so there would be a relatively high volume of activity on the day of the fieldwork. Four separate physical spaces were set out to match the research themes. The *connections* teams were located in Glasgow Central train station; the *transactions* teams were located in a shopping centre called The Buchanan Galleries; the *interactions* teams were located in Buchanan Street and the *conversations* team was located in George Square. Prior permission was sought to make observations in the train station and shopping centre locations. No permission was given to take photographs or record video in these locations, so these particular teams had to rely on note taking and sketches to record their observations.

4.1.4 Materials. Each group was given a plastic wallet containing: a briefing card, sketchbook, notebook and pencil.

4.1.5 Procedure. The study was split into 3 sessions:

- Fieldwork: a trip to Glasgow to carry out fieldwork observations.
- Dump & distil workshop: The group discussed the process of carrying out the observations and explored ways of presenting the gathered material.
- Affinity mapping workshop: Participants analyzed and synthesized the fieldwork notes and other types of data. The purpose was to arrive at a set of insights based on the observations that each team had made. Following on from this, the group tried to make overall sense of activities that had gone on that day across all of the original themes.

4.2 Fieldwork

The fieldwork activity began with a briefing in the city square. Participants were put into pairs, creating 7 teams and then the briefing packs were issued. The teams were given the option of assigning roles themselves but it was suggested that they split up the activity so that one person had the role of keeping notes and the other person the role of drawing sketches visually representing any observations. They were encouraged to keep communication flowing between each other whilst working on this activity.

The teams were asked to use the materials provided within the pack to help record their observations. The *interactions* and *conversations* teams, who were working outside, were also asked to use a digital camera to take photographs. The *transactions* and *connections* teams were working inside so had to rely solely on written notes and drawing sketches.

We scheduled 2 bursts of activity to take place on the day of the fieldwork. The first period lasted an hour and half, which allowed each of the teams to become acquainted with the research questions and familiarised with the locations where they were observing.

After the first burst of activity, the group gathered at a cafe to discuss how the fieldwork was going. This enabled them to talk about any problems that had occurred in the first pass of their separate spaces. Some issues that arose from this discussion were:

- Feeling overly visible whilst note taking and strategies for taking notes discretely.
- Having difficulty to decide what to focus on when there was too much happening at once to be able to record everything.
- Pretending to do something else other than observing so that the notes could be memorised and then recorded shortly after.

Once the group had finished their discussion, they split up once again into their individual teams and set off back to their designated spaces. This gave them another hour to make some further observations.

4.3 Dump and Distil Workshop

A week after the fieldwork, participants were asked to discuss how they managed the fieldwork process and how they interpreted the research questions that were set. They also answered questions related to the design of the study, the location and their use of the materials provided.

4.3.1 Approaches to Fieldwork. There was a mixture of approaches to making the observations in pairs. The *conversations* team decided to stay together throughout the fieldwork activity. This gave them the opportunity to have a continual dialog about what was being observed. Conversely, both the interactions teams separated whilst taking field notes and regrouped at set time intervals to discuss strategy about where to focus observations next (Fig. 1). A common approach was to take an initial walk through the space to become familiarised with it and then plan how to break their observations down into manageable chunks of activity. One of the *interactions* teams did this by splitting the timing of their observations equally for separate parts of the street so that they could systematically move down the street in 20-minute intervals.

All of the participants seemed to follow the instructions to allocate specific roles to each team member. There was commonly someone within each team who was stronger at recording things visually in comparison to taking written notes. Many of the teams switched roles at various points to try and gain a different perspective on the activity although, in relation to their drawing abilities, this meant that some team members were slightly outside of their comfort zone.

Where there were 2 teams working within the same space, they negotiated amongst themselves to try and eliminate the possibility of a crossover. For example, the 2 teams working on the *transactions* theme inside the shopping centre deliberately went to the highest and lowest floors so that more ground could be covered in the time available.

Fig. 1. One of the Interactions teams working out a strategy

A common insight was the groups finding it beneficial to have a higher point of view for making their observations from. The *connections* teams made use of the balconies that were available in the train station and the *interactions* teams used a set of steps outside a concert hall to gain a better vantage point. The teams felt that this provided a better overview of the activity that was occurring.

Many of the participants noted that they would like to return to the spaces again to go over some of the details that they might have missed the first time around. For example, checking on the sequence of events that a person who is interacting with a self-service kiosk has gone through. They also thought that returning at various times of day when the spaces are quieter or busier would be useful.

4.3.2 How to Share Data. The next step was to discuss how to go about sharing the fieldwork data between all of the teams. One of the participants suggested using a large physical map of the location to annotate their observations with over time. This might have been difficult for some of the interior spaces where maps were not so easily available. This also jumped too quickly to a physical space metaphor for understanding the data whereas at this stage, we needed to remain thinking about the data in a more abstracted way.

A secure virtual sharing space was set up for participants to place their fieldwork notes and images into in preparation for the final analysis workshop. This gave the participants an opportunity to see how other teams had approached the problem of how to keep easily readable field notes and the recording of other fieldwork data. The total amount of fieldwork data uploaded to the sharing space was 55 sketches, 175 photographs and 33 pages of notes. Each of the participants used a login name to sign into the shared virtual space so that the ownership of the content could easily be tracked. The fieldwork data was uploaded in a variety of formats ranging from single image files to full PowerPoint presentations. A standard format should have been agreed beforehand to make it easier to sort through the data in preparation for the workshop. Also, it would have been much easier to process the images if the individual instances of the image files were uploaded instead of, for example, embedding them within a PDF.

4.4 Affinity Mapping Workshop

In order to make sense of the vast amount of gathered fieldwork data, a final workshop session was held. The aim was to quickly reduce all data into a set of shared insights across all themes from the previous fieldwork activity. Given the time constraints, affinity mapping was selected as a suitable method.

The affinity mapping involved extracting notes and ideas from a larger piece of text or data then writing them onto sticky post-it notes. Once this was done, the next step was to try and organize the notes into sub groups to see if new themes emerge based around the particular notes that have been extracted. The workshop plan was to repeat this grouping process multiple times. Firstly, working within each of the small teams of 2 participants. Then working with both teams across each theme and finally everyone working across the overall themes.

4.4.1 Workshop Materials. There were a number of materials that were necessary to effectively carry out the workshop: a supply of post-it notes in 4 different colours; white index cards for recording theme headings; 8 sheets of A1 black paper; marker pens and drawing pins. Also, the day before the workshop, the group was asked to print out and bring along a hard copy of their fieldwork notes. A selection of the visual material that was uploaded to the shared folder was printed out and pinned onto the walls.

4.4.2 Field Data Tagging. Each of the teams was asked to make a pass through the opposite team's field notes working within the same theme. Each piece of extracted data was written onto a single post-it note and stuck onto a sheet of black paper in no particular order. The teams were asked to stick to one colour of post-its so where the note had come from it could be traced. The purpose of this exercise was to allow the team to offer a different perspective on someone else's field notes. This also meant that the field notes had to be written in a clear enough way so that other people could make use of the data.

4.4.3 Grouping Data. The participants returned to the post-it notes that had been recorded for them by the other team. They were then asked to read through the notes and start grouping these together as they deemed appropriate. Headings for each of the groups were written onto index cards and placed above them.

4.4.4 Grouping Themes. This task involved asking the 2 groups working on the same theme to bring together their notes to see where the commonalities were between them (Fig. 2). The two sheets of black paper were joined together so that there was a shared space for the groups to work within. This exercise allowed the teams to filter out duplicate sets of groupings. It also required the groups to explain the notes that they had written for each other previously and come to a shared understanding about the activity going on in each of the physical spaces they had observed.

The ways in which the groups manipulated the post-it notes in the space available was interesting. One of the groups moved all of their notes onto one side of the black sheets so that they had a clear space to organise them within. They could also be seen placing groups of notes at an angle on the paper as a way to differentiate them from the others. Once each of the teams had finished grouping their boards, they were photographed as a way of saving the layouts so that they could be referred to again later.

Fig. 2. The Transactions team's board showing their integrated data headings

4.4.5 Extracting Super Themes. At this stage, we had to zoom out of the original themes and try to arrive at an overall understanding of the activity that was taking place in the city centre on that particular day. The participants were given half an hour to look over the notes and themes that all the other teams had arrived at through the previous set of activities. All of the headings that were written onto the index cards were pulled off the wall and collected into a pile. The participants then gathered around a large table and all of the headings were laid out onto it. This allowed everyone to see where there might be some overall groupings that the headings could fit into. The headings that were duplicates and those that were quite vague, for example "Technology" were removed from the table.

Following a brief discussion amongst the group, it became apparent that there was a clear split between the interactions that people have with each other and their interactions with space and technology. The participants allocated separate areas of the table for each of these groupings and set about putting each of the headings into these spaces. The interactions that people have with each other grouping was given the title "Interpersonal Interactions" and 12 headings were placed under this. The interactions that people have with space and technology grouping was given the title "Ways of Interacting" and 26 headings were placed under this. A photograph of each of the groupings was also taken so that they could be referred to again later on.

4.4.6 Evaluating Super Themes. In order to find out if the themes that we had identified were appropriate, we had to refer back to our original post-it note field data. This would allow us to see how all of this data would fit into the new super themes that we had identified. The participants were instructed to split themselves into 2 large groups and divide the space in half down the middle of the table.

This provided the group with another new space to think about all of the gathered data that had emerged from an attempt to organise it under the super headings. The next stage of analysis for the participants would be to try to map all of the post-it notes along each of these newly identified axes. The workshop had already ran an hour over the original scheduled time at this point so was drawn to a close. Photographs of all the post-it layouts at the end of the session were taken to document how far the participants had progressed with analysing the data.

Fig. 3. Arranging the data into the super themes

The groups were then asked to pull all of the notes off the wall and place them under each of the super themes (Fig. 3). This enabled the groups to start from scratch again in relation to how they perceived the individual pieces of data. This introduced them to the idea of not being too precious about the original groupings they had set out and start thinking in new ways about organising the data.

This exercise enabled the group to see how valid each of the super headings were. The team working on the "Ways of Interacting" super theme split their groupings up again into 2 which were "Interactions with Technology" and "Interactions with Space". Also, some of the notes were split between the 2 super themes so were placed along a line down the middle between the spaces.

4.4.7 Identifying Axis. Due to the limitations of the tabletop space, the group quickly started to run out of room. In order to lay the data out more coherently, one of the participants suggested the need to arrange the data along different physical axis. Coincidentally, within the studio space there was a set of wooden exhibition stands. The participants used masking tape attached to them to mark out separate axis. These axes were 'person' to 'people' along the X-axis, 'tech' to 'non-tech' along the Y-axis and 'macro space' to 'micro space' along the Z-axis (Fig. 4).

Fig. 4. The axes the workshop participants identified

4.5 Navigating the Data

Once the workshop had been completed, there was a vast amount of data that would help to map out the problem. These were: the raw fieldwork notes & sketches; the digitised written up notes, photographs of the affinity boards at different stages of the workshop, the axis model, the unsorted post-it notes and index card themes.

When looking through this data, there seemed to be a regular need to zoom in and out of the different levels of material that was gathered. The different levels of information that are regularly accessed include:

– themes - top level providing colour-coded groupings of all themes created during the affinity mapping process.
– groupings - themes that have grown out of the arrangement of the individual tags.
– tags - recorded tags capturing any individual ideas based on an initial survey of the source material.
– source - the original field notes, sketches, images, sounds or video recorded whilst out in the field.

Fig. 5. A screenshot of the prototype tool

A web-based application was designed to prototype how these levels of information might be interacted with using a screen-based tool (see Fig. 5). The gathered material from the affinity-mapping workshop was used as example content to build this tool around. This prototype was only a sketch of an idea for how such a tool might operate. Future work requires enhancing the interface to allow zooming interaction for the content and adding the database backend to help manage the different types of media. In addition, user evaluation is needed.

4.6 Reflecting on the Method

The week after the affinity workshop, interviews with four of the participants were held to discuss their experience of working with the fieldwork data in a large group. The outcome of this discussion helped us to identify a number of points that could be explored further with the future iterations of the methodology. The issues we have identified were:

- Sharing context
- Descriptiveness of field data
- Physical access to the affinity boards
- Information overload
- Working across the different axis

4.6.1 Sharing Context. As the two teams worked together under each theme, there was a sense of familiarity with the field notes and being aware of where they had come from. An interviewee working on the *connections* theme noted that the opposite team had observed the same activity from another perspective so this helped to validate their own insights further. This had felt like they were “reading a guide book for a city they had visited before” so there was a clarity that could be gained from analyzing these observations.

It helped that the teams were working in the same place in the same time frame so they were able to crossover and relate perspectives. If they had to work with a group that had been located at different locations, it might have taken much longer to explain some of the meaning of what had been observed.

4.6.2 Descriptiveness of Field Data. When reading through other team's field notes, there was a reoccurring need to have set out a standard way for writing clear field notes so that others can make use of it more easily. Some sets of notes were written like a story in the format of long paragraphs. There is a need for a shared format that requires the fieldworker to indicate timings next to each of their pieces of data along with bullet points or sub-headings so that it is easier to derive key ideas from later on.

It was felt that the exchange of field notes was a useful exercise to gain a different perspective on the activity but important pieces of information were sometimes left out. Conversely, a fresh pair of eyes reading through the field notes occasionally pulled out new pieces of data that the original author might not have been able to spot.

When extracting the data onto the individual post-it notes, there was often difficulty with the level of detail that was recorded onto each note. For example, writing a single word that could have multiple meanings depending on how it is interpreted. There needs to be a pre-agreed level of detail for writing each note so that they can be easily shared and discussed.

4.6.3 Physical Access to the Affinity Boards. During the workshop, there was a contrast between the earlier tasks where the participants were working in small groups vertically on a wall and the later stages where the large group was working horizontally on a tabletop. The horizontal layout created a physical barrier of people so some had an advantage over others by it being easier to read the notes based on the position that they were in. This made some people quite dominant so they were blocking the view of others who were trying to read the notes. A solution to this problem would be to break down the large group across multiple tables so there is equal access and people are not continually jostling for the best position.

4.6.4 Information Overload. One of the interviewees stated that they had felt almost panicked at the huge amount of information that was laid out on the tabletop and it took real effort to re-engage with meaning behind each note. As there was so much to take in at once, the meaning of all the words had quickly started to disappear. A way to resolve this might be to start turning some of the notes into visual representations. This could help to make some of the concepts more approachable and free up thought process to concentrate on the ideas being represented instead of trying to always understand the meaning of the words.

During the final stages, there was often a lot of duplication in the data. It would be worthwhile to start stripping out some of the extraneous the notes and only work with the most important ones instead. This would help the key ideas come to the surface so they could be concentrated on and not get distracted by things of less importance.

4.6.5 Working Across Different Axis. At the end of the workshop, when the group had zoomed out of the final exercise, a lot of the information had been internalized at this point and lead to constructing a physical model with multiple dimensions of meaning. This started off with 2 dimensions but then it was realized that 3 dimensions would be necessary. The participants thought that it would be too difficult to think in 3 dimensions but they could quite easily work across 2 planes at a time so the analysis could be broken down and then rebuilt again afterwards. For example, with the person to people axis and tech to non-tech, these could be broken down into four separate quadrants and then the teams could work on these individually so they are not trying to think about too many things at once.

5 Field Studies 2 and 3

The method has been applied an additional two times both in education settings. These were at the University of Arts Zurich, Switzerland and again at the University of Dundee, Scotland with an alternative class. We will briefly describe the different conditions in comparison to the first field study and discuss how the methodology was refined based on our previous understanding.

5.1 Field Study 2

In Zurich, the same method was followed although the group who took part had less experience of using fieldwork techniques.

5.1.1 Participants. Eleven students on a Masters of Arts in Design participated in the study (8 male, 3 female). The participants were grouped into 4 separate teams with a theme assigned to each team.

5.1.2 Location. The town centre of Zurich, Switzerland was used as the location to carry out the fieldwork activity. The *transactions* team was located at the budget shopping centre ShopVille – RailCity; the *interactions* team was located at the upmarket shopping centre Globus Bellevue; the *conversations* team was based at the Masoala Rainforest Hall and the *connections* team was based at the Kunsthaus Zürich Museum for Modern Art. We did not have to seek permission to make observations in any of these locations and we were able to take photographs if required.

5.1.3 Materials. Each group was provided with a briefing card alone instead of the additional note and sketchbooks that were previously supplied. Participants were given additional hints on the briefing cards instructing them to assign roles for each team member between note taking, sketching and taking photographs. The cards also stated that the questions set out were only suggested and asked the participants to note down any additional questions that emerged as their observations progressed. A checklist was provided to the facilitator of the workshop to ensure that all the correct steps were followed for each phase of the methodology.

5.1.4 Fieldwork. The fieldwork phase of activity took place over two days, which allowed the teams to make a first attempt at observing and then return back to the studio to discuss what they had seen. This gave the group an opportunity to reflect on their progress early on and decide how to refine their fieldwork strategy further.

5.1.5 Results. Each of the teams produced a presentation of their findings following the final workshop phase of activity. This was a very useful additional task as it forced the group to consolidate the material they had gathered into a series of insights.

The *conversations* team even went as far as creating a short video featuring footage they had shot whilst out in the field. This video helped to convey their findings in an informative and accessible way.

5.2 Field Study 3

The method was applied an additional time with a different set of Design Ethnography students.

5.2.1 Participants. Fourteen students on the Masters of Design Ethnography participated in the study (9 male, 5 female). The participants were grouped into 4 separate teams and a theme was assigned to each team.

5.2.2 Location. The city centre of Edinburgh, Scotland was selected as the location to make the fieldwork observations. The teams were not assigned to any specific locations so they were given the freedom to select where would be most suitable for their particular theme. The only constraints for the participants were they could not go into any interior public spaces (railway station or shopping centre for example) as no prior permission had been sought.

5.2.3 Themes. The themes were altered slightly so the *connections* theme was replaced with a *consumers* theme and *conversations* was replaced with *environments*. A workshop prior to the fieldwork activity was also held that asked the teams to define their own questions based around each theme before going out into the field. Although this gave the teams greater ownership of the research questions, it made the task significantly more difficult.

5.2.4 Materials. As the teams had to define their own individual locations and themes prior to going out into the field, they were not provided with a briefing card as in the previous studies. However, each team was given a pocket-sized video camera that could be used to help document their observations.

5.2.5 Results. Following the fieldwork, the group was asked to create a short video to communicate their findings based on the observations they had analysed. It was useful to have this final piece of work to help summarise their analysis of all of the gathered material. It was also interesting to contrast how the group progressed without any tangible support materials when going out into the field. They seemed to be too ambitious with how much they were hoping to observe in the space of one day. However, the learning experience was improved greatly as without the support materials, the participants had to consider a lot more about how they planned their observations prior to the fieldwork activity.

6 Conclusion and Future Work

The method we have proposed sets out a new way of using multiple novice fieldworkers and so gaining multiple viewpoints when engaging with a design problem. We have been able to initially break down our problem into smaller parts and divide these up amongst a large group of novice fieldworkers. Such a group can explore the field setting and gradually co-create a rich understanding that cannot usually be gained by lone fieldworker observing in such a short time frame.

There is a clear need for tools to help collect, manage and analyse the fieldwork data when there is such a large group of people involved in the process. We have used an affinity mapping method with physical post-it noting sessions to do this but this has a number of limitations in relation to providing context around each of the pieces of data. Some existing tools, such as Prezi (prezi.com), offer some of the functionality required but lack the ability to manage multiple sources of material and deal with a variety of configurations of the gathered data.

There has been work investigating the use of tablet PCs [14] as a replacement for the existing lo-fidelity material in affinity diagramming. Another potential solution could be to use a multi-touch, multi-user technology platform that enables many people to work around the problem at the same time. This would help to alleviate some of the issues that we have identified and also enable sessions to be both co-located and take place between different locations.

Of course it should be noted that this method would not suit all types of problem. For instance, our approach to field data gathering would not be suitable for going into a small space such as a home with 14 people all observing the same thing. This would be uncomfortable for both the fieldworkers and subject! It would also completely disrupt the normal activity going on in such an intimate setting. Working across different types of public spaces provides enough room for a large number of fieldworkers to make their observations without continually overlapping with each other. Also, it is easier for the fieldworkers to make their observations without having too much effect on the normal behaviour of people within such a large space.

As our methodology has been greatly refined using the three field studies we have discussed in this paper, the next logical step is to apply it in anger within an industry context. It is here that the method could be of real benefit as the issues related to limited time and resources for carrying out lengthy field studies are much more pronounced. This is also the setting where any insights the methodology generates can be robustly evaluated to see how much value they can bring to the organization through gaining an understanding of their end users.

Acknowledgments. This research project is jointly funded by industry sponsor NCR and the Northern Research Partnership. We are grateful to the MSc Design Ethnography students who participated in the study. We thank the reviewers for their constructive comments.

References

1. Benkler, Y.: Coase's Penguin, or, Linux and the Nature of the Firm. The Yale Law Journal 112(3), 369–446 (2002)
2. Weber, S.: The success of open source. Harvard University Press, Cambridge (2004)
3. Howe, J.: The Rise of Crowdsourcing. Wired Magazine (2006) (June 14)
4. Salvador, T., Mateas, M.: CHI 1997 tutorial: Design Ethnography: Using Custom Ethnographic Techniques to Develop New Product Concepts (1997)

5. Millen, D.R.: Rapid ethnography: time deepening strategies for HCI field research. In: DIS 2000: Proceedings of the 3rd Conference on Designing Interactive Systems, pp. 280–286. ACM, New York (2000)
6. Myers, M.: Investigating information systems with ethnographic research. Communication of the AIS 2, 1–20 (1999)
7. Zeisel, J.: Inquiry by design: environment behavior neuroscience in architecture, interiors, landscape, and planning, Rev. ed. W.W. Norton & Company, New York (1979)
8. IDEO (Firm), IDEO method cards. IDEO, Palo Alto (2003)
9. Strauss, A.: Basics of qualitative research: techniques and procedures for developing grounded theory, 2nd edn. Sage Publications, Thousand Oaks (1998)
10. Kawakita, J.: KJ method: A scientific approach to problem solving. Kawakita Research Institute, Tokyo (1975)
11. Beyer, H.: Contextual design: defining customer-centered systems. Morgan Kaufmann, San Francisco (1998)
12. Pang, A.S.: Thinking Big: Large Media, Creativity, and Collaboration. Parsons Journal For Information Mapping III(1), 1–12 (2011)
13. Gault, P., et al.: Urban Traces – A Pilot Study into Public Space (2010) (unpublished document)
14. Mohamedally, D., Zaphiris, P.: Categorization Constructionist Assessment with Software-Based Affinity Diagramming. 25(1), 22–48 (2009)

Customer Experience Modeling: Designing Interactions for Service Systems

Jorge Teixeira[1], Lia Patrício[2], Nuno J. Nunes[1], and Leonel Nóbrega[1]

[1] Madeira Interactive Technologies Institute,
University of Madeira, Portugal
jorgegrenha@m-iti.org, {njn,lnobrega}@uma.pt
[2] College of Engineering, University of Porto, Portugal
lpatric@fe.up.pt

Abstract. Designers aspire to create engaging and desirable experiences. To that end they study users, aiming to better understand their preferences, ways of thinking and desired outcomes. In the service sector this task is more intricate as experiences encompass the whole customer journey, or the sequence of moments of interaction between customer and company. In services, one poorly designed interaction can severely compromise the overall experience. Despite experience holistic nature, current methods address its components separately, failing to provide an overall systematized picture. This paper presents Customer Experience Modeling, a novel multidisciplinary approach to systematize, represent and evaluate customer experiences to guide service and interaction design efforts. We illustrate this method with an application to a multimedia service provider built upon 17 interviews with service users.

Keywords: Customer Experience, User Experience, User Modeling, Service Design.

1 Introduction

Addressing experiences can be a daunting task. Experiences are unavoidable and ubiquitous, personal and subjective, and they are built from every contextual element in a given setting. Introduced by Pine and Gilmore [1] as an economic offering on its own, experience is nowadays discussed in interaction design [2, 3] and service design [4, 5]. Both consider experience a desirable feature and aspire to achieve engaging ones [5, 6]. For companies experience is consensually considered a source of sustainable competitive advantage [7, 8]. However, each field frames experience differently. Interaction Design considers users as the experience recipients and thus devotes its attention towards making products usable and agreeable to them [9]. This way user experience is viewed as an integrated offer of functional, or usability, benefits as well as emotional ones [7, 10]. In fact, some consider user experience as a better aim for Interaction Design and HCI than usability due to the narrow range of the latter [6, 10]. On the other hand, service design is concerned over customers, encompassing not only physical artifacts but the entire orchestration of service components such as physical environment, employees and service delivery process.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 136–143, 2011.

Whereas user experience necessarily involves an interaction, customer experience has been defined as the cognitive and affective outcomes of the customer's exposure or interaction with a company's people, processes, technologies, product, services, or other outputs [11]. Services do include products, or other interactive artifacts, but go far beyond them, being therefore more complex and difficult to design [12]. This study is purposively focused on customer experience as it aims to portray its holistic nature by encompassing every contact, direct and indirect with a service.

Existing methods already aim to handle the richness of experience data, albeit their focus is on its separate components. Personas depict representative users, service blueprints illustrate a service delivery process, use cases put forward a product, system or service intended functionalities, mood boards explore its emotional landscapes [13], to name a few. Experiences, however, are a combination of all aspects addressed by these methods and are perceived as a complex but unitary feeling [8]. As such, there is a lack of methods to handle experience as a holistic reality.

We introduce here Customer Experience Modeling (CEM), a method for interaction and service designers that supports the complexity inherent to service's customer experience. Leveraging contributions from well defined frameworks, concretely Human Activity Modeling [14], Goal-Oriented Analysis [15], and Multilevel Service Design [16] this method is able to systematize, represent and evaluate the several components that shape customer experiences .

2 Conceptual Foundations of Customer Experience Modeling

Models are abstract representations that weed out irrelevant details to focus on what is significant, explaining the operation and underlying concepts of systems that are too complex to be otherwise understood [17]. By using models to represent gathered knowledge we are able to synthesize and express it in a way that enhances communication between multiple stakeholders from different fields.

Models play an important role in interaction design as they help bridge the gap between problem and solution. According to the Analysis-Synthesis Bridge Model [18], the design process starts with observation and investigation of the current situation. Next, modeling is used to form a bridge between problem and solution, by helping interpret and systematize the understanding of the existing situation and explore new potential solutions. Finally, through an iterative process, idealized solutions are materialized into prototypes and ultimately finished forms.

CEM fits into this approach, offering a modeling framework that helps to represent and systematize gathered knowledge about customer experience and then shed this knowledge into interaction and service design efforts. However, capturing the complexity of experience required several contributions from different fields, namely Human Activity Modeling (HAM) [14] concepts and notation, Goal-Oriented Analysis' [15] softgoal concept and the three levels of experience from Multilevel Service Design (MSD) [16]. HAM gives the concepts and notation required to represent experience, the softgoals express desired characteristics to evaluate experience components and MSD provides a structure to systematize experience from an overall perspective to each single interaction.

2.1 Muldisciplinary Contributions

Human Activity Modeling (HAM) is a systematic approach to represent activities, developed to succinctly capture and represent them and their context to support practicing designers [14]. HAM provides the necessary conceptual backbone and notation to model interactions and its context, a first step towards customer experience improvement. Beyer and Holtzblatt consider context the backbone for organizing a customer-centric design process [19]. HAM's theoretical support lies on Activity Theory, a philosophical and cross-disciplinary framework that considers human activities as basic units of analysis and sees them mediated by artifacts [14, 20]. HAM also provides a representation, called Participation Map, of the participants and their relationships with each other and with the artifacts involved in an activity. This diagram is included in CEM to depict the contextual elements related with customer experience. However, HAM does not accommodate evaluation tools and is a system-centric approach, thus it needs to be adapted and integrated with other approaches.

CEM aims not only to represent and systematize customer experiences, but also to evaluate them in order to guide service and interaction design efforts. To accomplish this we need to elicit and portray user's activities, and also the goals behind them. Goal-Oriented Analysis' *softgoal* concept addresses experience subjective nature by accommodating partial, or intermediate states of satisfaction [15]. By introducing softgoals we can effectively evaluate each performed activity according to the desired requirements, thus detecting gaps or flaws susceptible to be addressed when designing service interactions. This way, it is possible to concentrate efforts on the most critical activities, modifying them or introducing new ones to enhance experiences. The concept of Softgoal is useful to represent experience factors, however they need to be complemented with models more focused on the service setting.

Multilevel Service Design (MSD) is a multidisciplinary method for designing services with embedded contributions from service development, interaction design and service design [16]. MSD proposes three levels of customer experience; value constellation experience, service experience and service encounter experience. Value constellation experience results from interactions between the customer and all service organizations needed to perform a given customer activity. Therefore it is not centered on a company's offering, but instead on the customer's activity and what services he uses to accomplish them. Service experience level shows how experience is shaped from different encounters with a single company, and which factors enable or inhibit the desired experience. Finally, service encounter experience level provides all relevant details for a single interaction, a specific encounter with the service, through a single channel. MSD offers a multilevel view of customer experience which already addresses some of its complexity. However, its representation still needs to evolve to provide a richer and concrete understanding of customer experience.

Customer experience can be represented, systematized and evaluated by integrating these multidisciplinary contributions. The next section presents CEM and shows how it was used to study a multimedia service provider customer's experience.

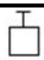
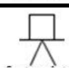
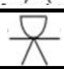
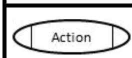
3 Customer Experience Modeling (CEM)

Representing the richness and complexity of customer experience requires an abstraction, a model with which the stakeholders can relate and use to communicate. Conceptually founded on an adapted customer-centric HAM [14], we followed the three levels of service design [16] and introduced the softgoal concept [15] to develop CEM, a multidisciplinary, multilevel approach, that captures and evaluates relevant features of customer experiences to support interaction and service designers.

3.1 Concepts, Notation and Structure

Since our objective is to guide interaction design and service design efforts from a broader contextual perspective to a single interaction we adopted the three hierarchical levels of MSD: value constellation experience, service experience and service encounter experience. Considering these three levels of service design enables us to trace each interaction from a single encounter to the overall value offering. HAM is the method's conceptual and notational foundation. However, we conducted a major adaptation in its perspective, as we regard it as customer-centric instead of system-centric. Therefore, the definition of each concept was altered to reflect this focus, as shown in Table 1.

Table 1. Adapted Notation for Customer Experience Modeling

Symbol	Name	Description
	Artifact	Any artifact employed within an activity [14].
	System Actor	Non human system (software or hardware) interacting with the customer.
	Role	Relationship between an actor and the customer.
	Actor	Activity participant interacting with the customer (or the customer himself).
	Softgoal	Condition in the world which the actor would like to achieve, but unlike in the concept of (hard-) goal, the criterion for the condition being achieved is not sharply defined a priori, and is subject to interpretation [21]
	Activity	Collection of actions or tasks undertaken for some purpose [14]
	Action	Action by a customer for some goal within an activity

To evaluate each activity and contextual component at different experience levels we introduced the softgoal concept [15, 21]. Figure 1 gives a generic view of CEM, showing how all these elements fit together.

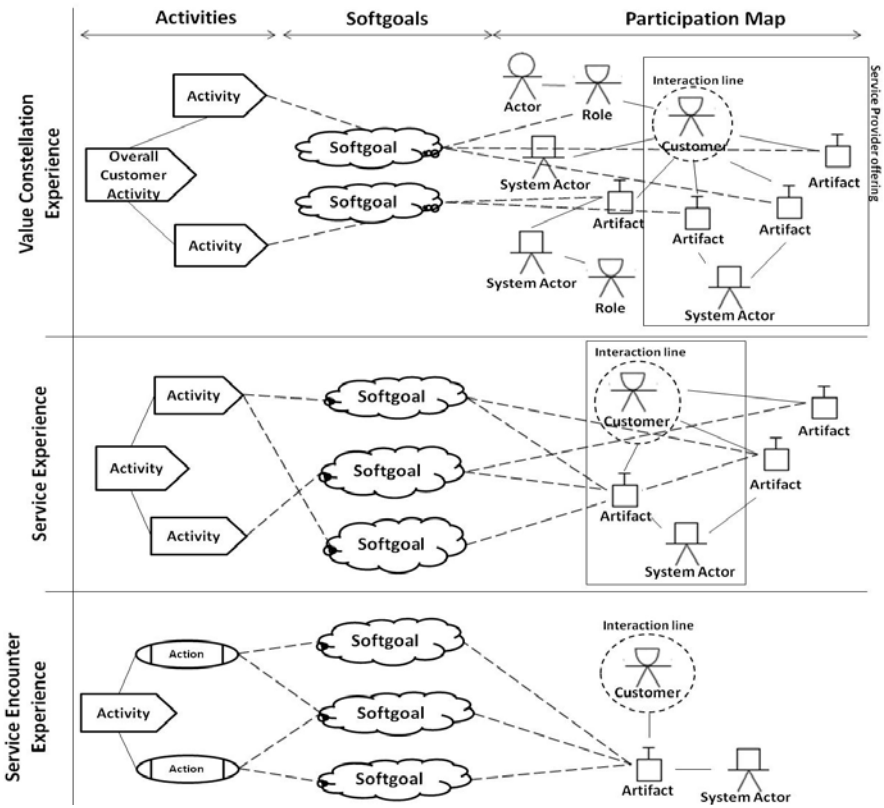


Fig. 1. Customer Experience Modeling generic view

3.2 Application of Customer Experience Modeling to a Multimedia Service Provider

CEM was put forth as part of a project to redesign the service offering of a large portuguese multimedia service provider. Its service offerings include cable TV, high-speed internet, phones and several video on-demand channels. The project involved a multidisciplinary team of business specialists, designers, and software engineers through an end-to-end effort, from studying customer experience, to implementing design chances in operational and system level. Customer experience data and the multidisciplinary team provided a fruitful ground for CEM's development. For this research, we interviewed seventeen customers and conducted a qualitative analysis of customer experiences, using grounded theory methodology [22]. The CEM for this multimedia service provider is shown in Figure 2. As we can see each level includes activities on the left, softgoals at the middle and, on the right, contextual elements represented by a participation map centered around the customer.

In the first level we are representing the value constellation experience, as such the participation map includes offerings, and consequent interactions, of many service providers. When crossing to the lower level, we select one of these service provider offerings and begin studying the customer's experience with it. In the bottom, the service encounter experience level is centered on a sole activity and the actions it comprises. At this level our focus is on very concrete interactions and, as before, softgoals and the participation map should reflect the concerns of the actions at hand.

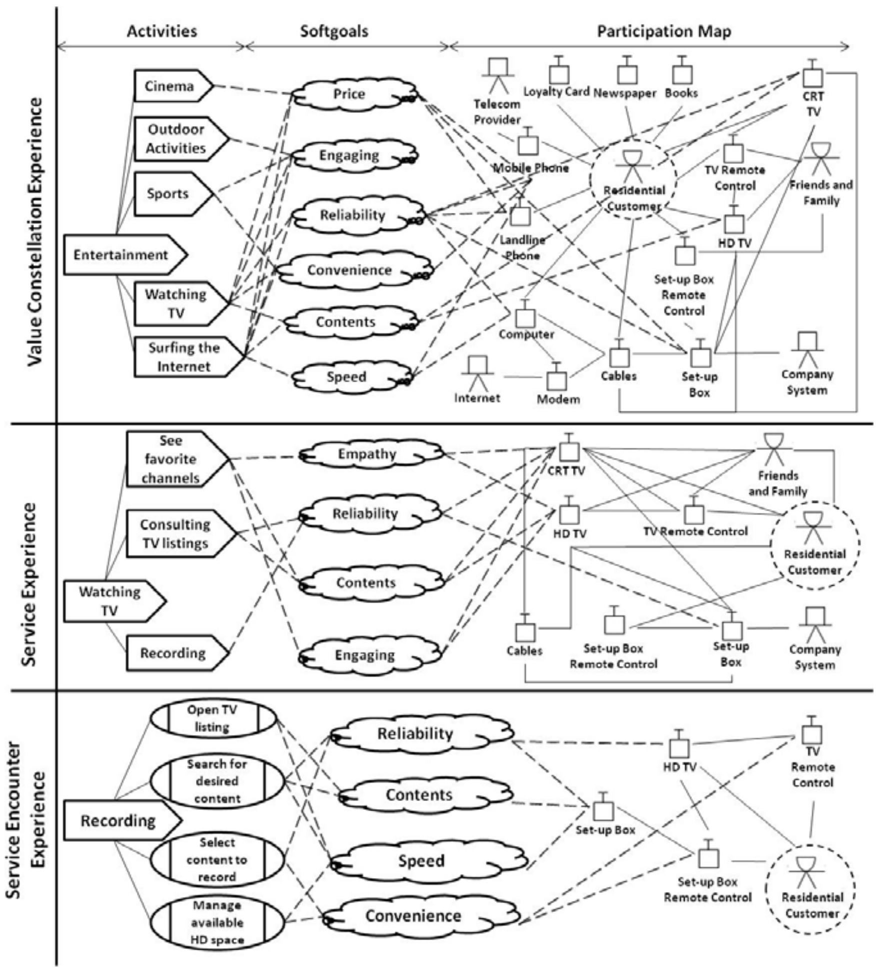


Fig. 2. Customer Experience Modeling for a Multimedia service provider

The participation map systematizes contextual elements related with activities within the same level. Therefore, the customer is surrounded by artifacts, systems and people with which he/she interacts. For that reason we have wrapped him within an interaction line, meaning that whenever he/she crosses that line he/she interacts with the immediate context around him. Every single interaction can be accounted for and related with softgoals for evaluation.

By analyzing the service provider's business model we selected the overall customer activity *Entertainment*. This overall customer activity is represented at the first level and is then divided into several more concrete activities, and connected to relevant softgoals. These softgoals represent the desired qualities, as expressed by the customers, of the activities and contextual elements. As we cross from one level to another, each activity partition reflects a new setting where the relative importance of softgoals is prone to change. We depict the most relevant softgoals for each level to give designers a granular and adapted view of experience for each activity. So, the first level shows that for *Entertainment*, customers consider activities like *Watching TV* or *Surfing the Internet*. In these activities they value *Price, Engagement, Reliability, Convenience, Contents* and *Speed*. These softgoals also characterize the contextual elements on the participation map such as *Computer* or *Set-Up Box*. Similar interpretation applies to the remainder levels.

In the end, we obtain a holistic representation of customer experience at its different levels, enabling interaction designers to trace each interaction from the corresponding overall activity. Considering Figure 2, the *Recording* activity must be reliable, fast, convenient and content-rich. Except for convenience, every other softgoals applies to the set-up box. By comparing with the level above we see that only customers with HD TV perform this activity. Also, for *Watching TV* customers concentrate the softgoals on TVs (CRT and HD), which is not supplied by the company. We conclude that a crucial artifact for customer experience isn't under company control. Both interaction and service designers can improve experience based on this knowledge.

4 Conclusion

CEM is a multidisciplinary method to represent, systematize and evaluate customer experience. It follows a customer-centric, multilevel approach to envision experience as seen from customer's eyes, with multiple service providers shaping the value constellation experience, and zooming to each service experience, and each service encounter experience. By merging contributions from different fields, such as interaction design, requirements engineering and service design. CEM supports design efforts and provides useful guidance to improve customer experiences.

References

[1] Pine II, B.J., Gilmore, J.H.: Welcome to the experience economy. Harvard Business Review 76, 97 (1998)
[2] Forlizzi, J., Battarbee, K.: Understanding Experience in Interactive Systems. Human-Computer Interaction Institute (2004)
[3] Marcus, A.: The Cult of Cute: The Challenge of User Experience Design. Interactions 9(6), 29–33 (2002)
[4] Patrício, L., et al.: Designing Multi-Interface Service Experiences: The Service Experience Blueprint. Journal of Service Research 10, 318–334 (2008)
[5] Zomerdijk, L., Voss, C.: Service Design for Experience-Centric Services. Journal of Service Research 13, 67–82 (2010)

[6] Law, E.L.C., et al.: Understanding, Scoping and Defining User experience: A Survey Approach. In: CHI -Conference-, vol. 2, pp. 719–728 (2009)
[7] Berry, L.L., et al.: Managing the Total Customer Experience. MIT Sloan Management Review 43, 85–90 (2002)
[8] Gentile, C., et al.: How to Sustain the Customer Experience: an overview of experience components that co-create value with the customer. European Management Journal 25, 395–410 (2007)
[9] Preece, J., et al.: Interaction design: beyond human-computer interaction. J. Wiley & Sons, New York (2002)
[10] Monk, A., et al.: Funology: Designing Enjoyment. In: CHI -Conference, pp. 924–926 (2002)
[11] Buttle, F.: Customer Relationship Management: Concepts and Technologies. Elsevier Ltd., Amsterdam (2009)
[12] Norman, D.A.: Living with complexity. MIT Press, Cambridge (2011)
[13] Saffer, D.: Designing for interaction. Pearson Education (distrib.), Berkeley (2010)
[14] Constantine, L.: Human Activity Modeling: Toward A Pragmatic Integration of Activity Theory and Usage-Centered Design. In: Human-Centered Software Engineering: Software Engineering Models, Patterns and Architectures for HCI, pp. 27–51. Springer, London (2009)
[15] Mylopoulos, J., et al.: From Object-Oriented to Goal-Oriented Requirements Analysis. Communications of the ACM 42, 31 (1999)
[16] Patrício, L., et al.: Multilevel Service Design: From Customer Value Constellation to Service Experience Blueprinting. Journal of Service Research 14, 180–200 (2011)
[17] Ludolph, F.: Model-Based User Interface Design: Successive Transformations of a Task/Object Model. In: Wood, L.E. (ed.) User Interface Design: Bridging the Gap from User Requirements to Design. CRC Press, Boca Raton (1998)
[18] Dubberly, H., et al.: The Analysis-Synthesis Bridge Model. Interactions 15, 57–61 (2008)
[19] Beyer, H., Holtzblatt, K.: Contextual design: defining customer-centered systems. Morgan Kaufmann, San Francisco (1998)
[20] Kaptelinin, V., Nardi, B.A.: Acting with technology: activity theory and interaction design. MIT Press, Cambridge (2006)
[21] Yu, E.S.K.: Modeling Strategic Relationships for Process Reengineering. In: Yu, E.S.K., et al. (eds.) Social Modeling for Requirements Engineering, pp. 11–152. MIT Press, Cambridge (2011)
[22] Charmaz, K.: Constructing grounded theory: a practical guide through qualitative analysis. Sage Publications, London (2006)

User Experience Research in the Semiconductor Factory: A Contradiction?

Marianna Obrist, Wolfgang Reitberger,
Daniela Wurhofer, Florian Förster, and Manfred Tscheligi

Christian Doppler Laboratory for "Contextual Interfaces",
HCI & Usability Unit, ICT&S Center, University of Salzburg,
Sigmund-Haffner-Gasse 18, 5020 Salzburg, Austria
firstname.lastname@sbg.ac.at

Abstract. No doubt, user experience (UX) has become of high relevance within the HCI community. Within this paper, we present initial results from a qualitative study on UX in the factory context, more precisely in a semiconductor factory. We highlight the challenges of performing UX research in this particular context and the usefulness of probes for collecting feedback from operators in the factory context within in a limited timespan. The results provide an initial rich description of the operator's everyday experiences in a semiconductor factory. From a designer's point of view, this allows for a more empathic interaction design informed by a subjective operator perspective.

Keywords: user experience, factory context, probing, user study.

1 Introduction and Motivation

So far, only little work focusing on the interaction and user experience (UX) in factories has been done. However, UX in a factory context is crucial. For example, cooperation and reliability of workers – often mediated through technology – are essential for the course of actions in the factory. Human errors resulting from poorly designed interfaces lead to high costs and even defective end products. Additionally, emotions – although not primarily associated with factory work – might influence the cooperation between factory workers as well as efficiency and reliability. Thus, UX and the related methodological challenges in the factory context represent an important research topic. Within this paper we introduce how we dealt with the challenges and limitations in a semiconductor factory when conducting an experience-focused user study. We will show how we successfully adopted the probing method for investigating UX in this particular context.

First of all, to create awareness and foster appreciation of UX research in the factory, a workshop with members of the factory management was held. The research goals and the benefits of a qualitative study were discussed. Initial doubts on whether a creative method like probing would work for a group of fairly unskilled workers, were dispelled. At the end, the method proved to not only work well but also lead to relevant findings. The adapted probing method engaged the operators (i.e. workers) and motivated them to give a rich account of various aspects of their everyday work

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 144–151, 2011.

experience, including open input about perceived negative issues such as stress. The deeper understanding of the subjective perception that operators have of their work context has the potential to better inform designers and increase their empathy for this particular user group.

2 UX Research in a Semiconductor Factory Context

In the following we describe the potential of UX research for the factory context considering the particularities of a semiconductor manufacturing plant and the resulting methodological challenges.

2.1 Relevance of User Experience Research

Following Hassenzahl and Tractinsky [7], UX is a consequence of a user's internal state (e.g., predispositions, expectations, needs, motivation), the characteristics of the system (e.g., complexity, usability, functionality, etc.) and the context where the interaction occurs (e.g., organisational or social setting). UX itself can be understood as a dynamic, complex and very subjective phenomenon [10]. Moreover, according to Alben [1], UX comprises all 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 are using it, how well it serves their purposes, and how well it fits into the entire context in which they are using it. This definition particularly points out the relevance of UX research in all areas where humans interact with a system, thus also in a semiconductor factory context, where people are increasingly confronted with new technologies and user interfaces to complete their tasks.

2.2 Understanding the Context and Challenges

The main challenge for semiconductor manufacturing plants is the coordination of many operators working on different machines to guarantee an efficient and error free production process. As soon as we talk about factory ergonomics like usable, safe, and comfortable interfaces, we are also addressing aspects of UX. Despite the obvious relevance of UX in the factory, little research has dealt with this issue. This could be partly rooted in the difficulties and limitations such a context involves (see below) or due to the fact that the investigation of UX might lead to competitive advantages, and are thus not published for a greater audience.

One particular challenge of this context is the "clean room" with its special environmental conditions. In a clean room it is necessary to wear a special outfit covering the whole body and to use a mask. Traditional study materials, such as paper and pen, recorders, etc. are not allowed to take into the clean room. Additionally, shift-production circles – the manufacturing plant operates 24/7 with day and night working shifts – can also impede a field study in this context. Consequently, doing a field study on UX in a semiconductor context often requires modifications and adaptations of traditional methods.

For decades scientists were and are still being occupied to investigate factory work from a point of view of classical or social psychology [3]. The factory context is a wide area of research within the field of ergonomics, also focusing on technologies and its impacts on workers, especially occupational safety and health issues [17]. However, little research has been done to investigate UX issues within a factory context so far (e.g. [8]). Few materials can be found that put factory workers into the focal point of attention when designing interfaces. In the beginning of the 90s an “IEE Colloquium on HCI: Issues for the Factory” was held in London, dealing with psychological basis for computer system design, operator support systems and industrial inspection [13], involving an article on the design of operator-computer interfaces that give operators an accurate picture – acquisition and display of data – and allow the control of the production process. Going a step further and trying to understand how operators feel about and experience the interaction with diverse interfaces, is still rare with respect to the increasing relevance of UX research.

2.3 Using Probing to Investigate UX in the Factory Context

To investigate HCI issues within the factory context, some studies have been conducted using different methods. Reul [16] reports on the improvement of usability of control software for industrial test systems. By using a contextual inquiry he could better understand the users' tasks and identify wishes of different user groups. For the redesign of the control room of a large factory Sagar [18] used contextual inquiry and observation as methods to gather information and understand the different tasks of the operators. A state of the art on control room design for industrial plants (e.g., power plants) has been published by the Technical Research Institute of Sweden [4]. Other reports cover usability issues in complex industrial control systems [14] especially focusing on nuclear power plants. There is no research as far as we are aware of, using more creative and engaging methods, such as probing, in the factory context.

The probing approach is used to help designers understand and specify the context of use and to support them to produce design solutions [12]. Probes/probe packages including creative materials, such as diaries, cameras, postcards, etc., are in general used as a means to offer inspiration for the designers or as a means to gather qualitative user information [2]. Originally developed by Gaver et al. [5], cultural probe packages contained open-ended questions, provocative and oblique tasks to support early participant engagement within a design process. Probes record a point of view, in the moment experience by making particular action, places, objects, people, etc. visible as well as motivations, wishes, desires, emotions, and intentions [6][9].

The central element of probing is that users are given materials to describe their life or their work in a self-motivated way within the respective context. Users thus reveal new insights into practices and thoughts that cannot be revealed by classical approaches like questionnaires, interviews and observation. Probing offers an unobtrusive way to investigate people's attitudes and goes beyond reflective opinion gathering, as it fosters creativity. As Lucero et al. [11] point out, probes can – among other goals – facilitate looking into participants' daily practices and finding inspiration for new concepts.

Conducting UX research in such a particular context as the factory, using a design inspired method, was considered relevant for collecting feedback from the workers and providing valuable insights on further improvements of the work processes.

3 Field Study Set Up

The probing method described in the previous section had to be adapted for studying the operators' experiences in the factory context due to the nature of the study environment and the limitations of the clean room. In the following we will give an overview of the target objectives and present the detailed study set up.

3.1 Goals and Objectives

The following objectives were defined for the conducted study (combining goals for UX research for this particular context and methodological insights):

(1) Investigate user experience of workers within the factory context.

(2) Apply a creative approach, inspired by probing, which is applicable for this context in order to investigate workers' experiences.

The first objective aimed at understanding UX within the factory environment, mainly the clean room. Thereby, we also tried to find out more about factors influencing UX, including aspects like the social network and interactions of the operators. By means of a creative stimuli approach, the study participants were asked to express themselves and explain their viewpoints concerning work with different devices and user interfaces within the clean room (e.g., see [13]). The second objective deals with the issue of how to adapt the probing method in a factory context to get insights on UX. For example, the time constraints of workers heavily influenced the creation of the study set up and the probing materials. Apart from the fact, that operators don't have the time to fill in probe packages during their work processes and the time slots before and after their individual shifts are too short to get reasonable results, we had to modify our set up to fit the contextual circumstances. Especially, the clean room poses certain limitations to classic probing materials, e.g. no normal paper and pen or cameras as material are allowed.

3.2 Study Set Up and Materials

After considering different procedural possibilities with the management (like handing out material to the operators for usage at their home, providing probing sets in the cafeteria, etc.) we decided to use one of the regularly scheduled instruction sessions to distribute the probe packages. These sessions are held at fixed intervals (approx. once a month), usually one hour before the normal shift starts. They are used to teach operators new skills or instruct them about changes/updates of the systems they are working with. These training sessions seemed ideal for carrying out our probing study. Although limited in time, this set up had the advantages, that the workers are already in the work context and do not need to use their spare time to do the probing (which is critical in this context). The study has been conducted on 12th of October 2010 at a one-hour instruction session at the manufacturing plant. Participants (N=55)[1] were operators working at the clean room. Two researchers explained the tasks and the aims of the study to the participants. Furthermore they

[1] No socio-demographic data were collected as it was not seen as relevant for this initial study and in order to keep the anonymity of the operators as requested by the management.

gave a short overview of the different probing sets available. The probing material consisted of a variety of different materials, each focusing on a specific topic or task. The probing material included eight probing sets and was divided into two categories.

(1) Probes focusing on UX of workers and working conditions influencing the UX
This category comprised six probing sets. The first three probing sets aimed at investigating current factors that account for the UX of workers within the factory context. They included questions about current mistakes that occur regularly, namely disturbing issues (see Fig. 1, "Things that bother me") and asked the participants to express wishes to facilitate work (see Fig. 1, "The magic fairy").

Fig. 1. Three examples of probes used in the factory context

The other three probing sets were designed to reveal more insights into daily work practices of the workers and factors that influence the UX. The first set wanted the workers to explain what constitutes a perfect workday (see Fig. 1, "The perfect working day"). The second set asked workers to describe their "neighbourhood", thus investigating to what extend the workers collaborate with others and how they experience the collaboration. The third probing material asked workers to draw their "level of stress for a typical day" on a time-graph, and explain highs and lows.

(2) Probes dealing with feedback for existing and planned interfaces
Aiming at involving users into the design process of interfaces we developed two probing sets that asked for their ideas and opinions on interfaces. In the first set they had to rearrange an existing interface according to their preferences grounded in their work practises. The second set was similar, but focused an interface that was currently developed. As we focus on UX within this paper, we do not present these two probes. Each of these probing sets contained a separate, quick description of how to use the material, including pens, scissors and glue. Every participant was able to choose one or more sets of probing material, depending on his/her interests. In the following the findings on UX on the factory context are presented.

4 Relevant Findings on UX

The collected probes (N=75 in total, of which 32 were relevant for UX) were transcribed as well as scanned in order not to lose any not-textual information. They were analysed using NVivo, a software package to support qualitative research (http://www.qsrinternational.com). We conducted a qualitative content analysis supported by statistical measures such as word frequency counts and rankings. Based on this analysis of the probing materials, insights on the UX factors stress, usability/ergonomics, and emotion could be gained. Moreover, social aspects turned out as a major influencing factor on UX in the factory. These factors (stress, usability/ergonomics, emotion, social aspects) can be seen as important in the factory context, as they were mentioned by the workers themselves, without giving them any triggers (open questions of the probes “Perfect Working Day”, „Things that bother me“, „Magic Fairy“). Furthermore, specific insights on stress were revealed by the probe “On Stress”. Table 1 gives an overview on the probes revealing insights on specific UX and influencing factors.

Table 1. Factors revealed from different Probes

	Stress	Usability/ Ergonomics	Emotion	Social Aspects
“Perfect Working Day” (N=10)	x	x	x	
„Things that bother me“ (N=9)		x		x
„Magic Fairy“ (N=3)				x
“On Stress” (N=10)	x			

Stress: Stress turned out to be of high relevance for the workers’ experience in the factory context, as it was mentioned often in different probes. When asking the participants how a perfect working day is characterized, stress was mentioned most often (in 80% of the probes). In particular, a perfect working day is mostly characterized by the absence of negative characteristics like stress or fear. Reasons mostly mentioned for stress are the end of a shift and shift change. At this time, the remaining work must be organized according to the established informal rules to avoid interruptions in the working process and enable a smooth takeover by the next shift group. Most of the time during a shift is experienced as stressful (67% on average).

Usability/Ergonomics: The workers perceive usability and ergonomics as important. This factor is closely connected to the factor stress, as usability problems are often mentioned as a source of stress. When asking workers what bothers them in their work, they often mention usability and work organization (21% of mentions). For example, workers bother when the machines are difficult to handle, or they do not like when there is not enough space for acting. Vice versa, a perfect working day is characterized by the prevention of errors and (bad) equipment (i.e. usability) (24% of mentions). Thus, the absence of usability problems is the second frequently mentioned reason for a perfect working day surpassed only by the absence of stress.

Emotion: At first glance, emotion does not seem to play important role in factory context. Due to the rather open design of some cards, we were able to identify emotion as a relevant aspect of UX in the factory. When analysing the probes in detail, it turned out that negative emotions like fear and anger are mentioned in relation to work. Positive emotions cannot be found in the probes. As most of our questions focused on drawbacks and deficiencies at work, the occurrence of negative emotions can be partially explained by the design of the probes.

Social Aspects: Social aspects probably represent the most important influencing factor on UX in the factory context. Probes concerning the social aspect show that workers desire more positive feedback from the boss or superiors, appreciate the reliability of colleagues, and demand equal treatment for all operators. When participants have three wishes, they mostly mention social aspects to be improved (e.g. better cooperation between colleagues). Asking workers for things that bother them at work, social aspects are often mentioned (e.g., staff, equal treatment, social contact; mentioned in 48% of the probes).

5 Discussion and Conclusions

No doubt, user experience (UX) has become of high relevance within the HCI community. Especially when the focus of HCI shifted from the work context to the more private and home context, the UX concept appeared. Designing for a fulfilling and enjoyable interaction with interactive systems has become a main goal for interaction designers. However, when it comes back to the work context, what significance can or should UX have in this context? Our study results showed that UX factors, such as emotions, stress, and social aspects, play an important role in the daily working routine of operators in a semiconductor factory. Investigating the workers experiences in this particular context was one of the main goals of our presented study using creative probes as stimuli. The results showed that the absence of stress significantly contributes to a perfect working day and that the end of a shift is experienced as the most stressful part of the daily working routine, as this is the most critical point of time for a fluent working process.

We further aimed at developing a creative approach applicable for the factory context. Probes can provide a good input for UX in the factory context. Open and unstructured probing cards (e.g., “Make a wish”, “Things that bother me”) give hints on which UX or influencing factors are important for workers. Thus, first insights on relevant UX factors can be gained. In a second step, insights on specific factors could be extended by triggering questions on specific factors (e.g., “On Stress”). Finally, probing has proven its relevance for investigating UX in the factory context. The adapted probing method turned out to be an appropriate method for providing easy and fast feedback of workers taking into consideration the time constraints and contextual circumstances. Thus, we developed an increased understanding of individuals’ as well as on the groups of operators’ user experience. As a next step we plan to link UX research back to user interface design and specific interaction problems in order to improve workers’ experience in their daily work routines.

Acknowledgments. The financial support by the Federal Ministry of Economy, Family and Youth and the National Foundation for Research, Technology and Development is gratefully acknowledged (Christian Doppler Laboratory for "Contextual Interfaces". Special thanks go to Florian Pöhr for his effort in the study.

References

1. Alben, L.: Quality of experience: defining the criteria for effective interaction design. Interactions 3(3), 11–15 (1996)
2. Arhippainen, L., Tahti, M.: Empirical evaluation of user experience in two adaptive mobile application prototypes. In: Proc. of the 2nd International Conference on Mobile and Ubiquitous Multimedia, pp. 27–34 (2003)
3. Brown, J.A.C.: The social psychology of industry: Human relations in the factory. Penguin Books, Harmondsworth (1954)
4. Ericson, M., Strandén, L., Emardson, R., Pendrill, L.F.: prestudy - State of the art survey of control rooms. SP Technical Research Institute of Sweden (2008)
5. Gaver, B., Dunne, T., Pacenti, E.: Design: Cultural probes. Interactions 6(1), 21–29 (1999)
6. Graham, C., Rouncefield, M., Gibbs, M., Vetere, F., Cheverst, K.: How probes work. In: Proc. of the 19th Australasian Conference on Computer-Human Interaction, pp. 29–37. ACM, NY (2007)
7. Hassenzahl, M., Tractinsky, N.: User experience - a research agenda. Behaviour & Information Technology 25(2), 91–97 (2006)
8. Hassenzahl, M., Platz, A., Burmester, M., Lehner, K.: Hedonic and ergonomic quality aspects determine a software's appeal. In: Proc. CHI 2000, pp. 201–208. ACM, NY (2000)
9. Hemmings, T., Crabtree, A., Rodden, T., Clarke, K., Rouncefield, M.: Probing the probes. In: Proc. of the Participatory Design Conference, pp. 23–25 (2002)
10. Law, E.L.C., van Schaik, P.: Editorial: Modelling user experience - an agenda for research and practice. Interacting with Computers 22(5), 313–322 (2010)
11. Lucero, A., Lashina, T., Diederiks, E., Mattelmaki, T.: How probes inform and influence the design process. In: Proc. DPPI 2007, pp. 377–391. ACM, NY (2007)
12. Mattelmaki, T.: Design probes. University of Art and Design Helsinki, PhD thesis (2006)
13. Meech, J.: Intelligent systems for operator support. IEE Colloquium on HCI: Issues for the Factory, 2/1–2/6 (1991)
14. Meschtscherjakov, A., Reitberger, W., Pöhr, F., Tscheligi, M.: The operator guide: An ambient persuasive interface in the factory. In: Proc. of the First International Joint Conference on Ambient Intelligence 2010. LNCS, pp. 117–126. Springer, Heidelberg (2010)
15. Norros, L., Savioja, P.: Usability Evaluation of Complex Systems - A literature review. STUK - Radiation and Nuclear Safety Authority, Finland (2004)
16. Reul, M.: Bringing usability to industrial control systems. In: Proc. of CHI EA 2009, pp. 3335–3340. ACM, NY (2009)
17. Robinson, M.A.: Work sampling: Methodological advances and new applications. Human Factors and Ergonomics in Manufacturing & Service Industries 20, 42–60 (2010)
18. Sagar, M.: User interface prototyping as seen from three viewpoints in the context of a supervision system. In: Proc. of the 15eme Conference Francophone sur l'Interaction Homme-Machine, pp. 190–197. ACM, New York (2003)

Client's Temporal Trajectory in Child Protection: Piecing Information Together in a Client Information System

Saila Huuskonen and Pertti Vakkari

School of Information Sciences,
33014 University of Tampere, Finland
{saila.huuskonen,pertti.vakkari}@uta.fi

Abstract. Our study focuses on the information needed for overviews by social workers in child protection, and how information is presented in client information systems (CIS). Data consists of semi-structured interviews and social workers' observations while they used CIS in their daily work. The analysis was structured by the concept of temporal trajectory. We identified three major interconnected information strands: concern, child's private life and institutional actions. Their temporal lengths and aspects are disparate. CIS offers modest temporal overviews of clients' cases. Representing information strands as timelines on the interface would provide better overviews.

Keywords: client information system, child protection, temporality, work tasks.

1 Introduction

Child protection is a part of statutory social work. Its objective is to support children and their families in problematic life situations. Primarily work is done by offering in-home services. In child protection services, a client information system (CIS) is used to support social workers' work. It has a crucial role conveying information about a client's case. Information is recorded and used in CIS during a whole client process as part of daily work tasks [8]. The representation of the recorded information as a case overview in CIS is in the focus of our study.

Information is collected mostly in face to face interaction with different parties, and it is afterwards selectively recorded in CIS. Information in CIS can be plentiful especially in long and active client processes. It gets demanding to piece together an information puzzle to form an extensive overview of a case. Thus, fast and effective information representation in CIS would be essential so that social workers could rapidly make sense of how a client's case has evolved over a period of time. The overview available at-a-glance has value in several work tasks [8]. This is the case, for example when a new worker takes over a case or an old timer with heavy case loads tries to summon up a memory of what the client's story is**.** Obviously, the effective overview in CIS constitutes two intertwined things: the essential information that is well documented and CIS' ability to present it in a concise manner.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 152–169, 2011.

Social workers' work orientation is holistic. This means understanding a person as a whole and as a part of social context [9]. The work requires a wide perspective about a client situation. However, the current systems do not seem to offer appropriate tools to present a client's situation fully. Systems are criticized, splitting holistic case information into pieces [7], [16], [33] and not fulfilling the task they were originally meant for [3]. Thus, a misfit between the needs of social workers and the way CIS present information seems to be obvious.

Despite the strong criticism there is a lack of studies, which analyze social workers' perception of holistic information needed in their work tasks, and how CIS supports representing and using this information. By exploring in detail what kind of overviews of client cases are needed we can create results for developing CIS to better fit to social workers' tasks.

We approach the child protection information from the temporal perspective. The importance of temporal perspective in research has been recognized in information studies [24], [26], social work research [19], [31] and CSCW/HCI [20]. Our research is situated at the intersection of these three disciplines.

Our research has three objectives. Firstly, we examine what kind of information according to social workers belongs to a comprehensive temporal overview of a client case in child protection's in-home services. Secondly, we give an account of how this information is currently presented in CIS in our research sites. Thirdly, we relate a need for 'overviewable' temporal information to work tasks. Temporal trajectory [20] [28] is used as an analytical concept. This concept offers an analytical lens to look at social workers' work with a client's case in temporal context.

Our study is based on field work realized in in-home services in child protection in Finland. In-home services are offered to a child and his family based on child welfare assessment. The services are defined by legislation. The field work consists of social workers' interviews and observation in real working situations with CIS.

The paper is structured as follows. In the first section, we describe the conceptual framework and review the previous literature. Next, we present the research setting including the research site, data collection and analysis methods. Then we introduce the results starting by reviewing social workers' perception of information that belongs to a client's temporal trajectory. Then, we analyze how the information with temporal perspective is presented in the CIS of our research sites. The last part of the result section looks at a temporal trajectory and CIS as a part of social workers' work tasks. We conclude with binding our three research objectives together, and outlining some design ideas.

2 Conceptualization of Research

Time and temporality both have a central place in social work practice. Many social work activities, the notion of intervention and change and social workers' sequential narrative accounts of cases have an aspect of temporality [31].

Temporality is embedded in different types of processes in social work. The processes have been defined from the perspective of work as the series of encounters and activities during the professional services [5]. The processes can be approached from the perspective of an individual client, too. Then a process is seen through human growth and changing circumstances [9]. The concept of trajectory originates from the studies that described work in medical settings. The central idea of the trajectory was to depict the organization of work associated with the course of the illness of a patient [28].The generalized concept of trajectory is understood as the evolvement of a phenomenon over time plus action, interaction between multiple actors and unanticipated contingencies related to phenomenon [27].

A further developed concept, temporal trajectory, is defined as a structured timeline that consists of activities, events and occurrences over time. The temporal orientation covers the past, present and expectation of future activities. This wide perspective makes it possible to see the relationship of activities and to take a look at patterns of former activities and anticipate the progress of a case. [20]

The temporal trajectory embodies similar elements as narrative accounts, traditionally valued in social work. Narrative has thematical development; it shows causality and it tells the story with a beginning, middle and end [1] just like temporal trajectory presents evolvement of a case over time.

We complement our analysis with three aspects of temporality: episode, interval and eon. Episode means a short period of time; interval a long period of time with distinct starting and ending; eon is a long continuous period of time [26]. The aspects are used to give a temporal shape to information in the trajectory.

3 Related Research

The core critical arguments in social work towards CIS concern the way of presenting information. The CIS has been seen as a threat to holistic and narrative information [16]. Reasons for the misfit might be that information is forced into a certain format [22] with too much focus on facts [14] and by this way being irrelevant to practitioners.

Recently, there have been plenty of studies around the British Every Child Matters Reform and its technological solutions. The main result of the studies is that this new technology does not optimally support the work practice. Integrated Children System does not offer a complete picture of a client's case because of a lack of contextualized narrative [33] and a lack of a common family file [7]. The research realized more than ten years earlier among Norwegian social workers ended up with similar results: it was impossible to pull together the full client history to view it at a quick glance since information was as pieces on the computer screen [9].

Physicians shared similar problems in gaining an overview from electronic heath records. Navigation between different screens lost the overview [2]. The studies seem to indicate that different professions share similar needs and similar problems.

Studies on reading the case files indicate that there is need for fast access to information. In CIS information is skimmed through with the aim of locating the most essential parts of it. Rapid access for information representing case overviews was valued, in particular. [8] The similar type of skim reading was recognized already in the time of paper documents both in social work [12] and in the medical context [15].

A structural client's assessment form (CAF), also part of the British Reform, splits information into fragments [25], disrupts the temporal and narrative display of information [32] and limits the expression of concern by highlighting a more robust assessment of needs [7]. The CAF form was used more in parts that offered possibilities to produce traditional narrative information [17]. Findings are alike in the study realized in Israel in the juvenile home context, where the structured assessment form was considered cumbersome and incapable to capture the client's unique information [23].

The temporal aspect of information and its implications for system design have been under study in computer supported co-operative work. The studies have been realized in health care. Health care staff valued seeing information organized chronologically because in that way they were able to situate the patient within the temporal trajectory and gain better understanding of a current situation [20]. The need for retrospective or prospective information depends on the work task at hand [21]. The study concerning chronic patients' illness trajectories highlighted the value of a good overview of a patient and as well an evolvement of medical information over time [13].

The previous studies point out a need for a chronological, time-wise approach to structure information, and the need for an overview of a client's situation available at a glance.

4 Research Design

4.1 Research Questions

We are interested in a triangle that constitutes information, its temporal presentation in CIS and how they intertwine in social workers' work tasks in child protection. The precise research questions are:

- What information strands belong to a client's temporal trajectory as perceived by social workers?
- How is a temporal trajectory presented in CIS in our research sites?
- How do social workers manage a temporal trajectory in CIS?

4.2 Study Environment

The field work took place in three Finnish municipal social service organizations offering in-home services in child protection. The in-home services defined by the Child Welfare Act (417/2007) include, e.g. support persons and families, financial help and family workers' visits at home. The data was collected in the year 2008.

In city A, the number of front line social workers in child protection in-home services was ten, in city B nine and in city C 35. Outside of office hours an emergency department was responsible for urgent situations. Staff turnover was relatively high in the research sites. Often several social workers took care of a client's matters over time.

A social worker was in charge of a varying number of children. Most of the time a co-worker was also named for a case. Among interviewed social workers the caseload was heavy, varying from about thirty up to over a hundred cases. Social workers' task was to manage a client’s case, e.g. by assessing need for the services, keeping in touch with a child and his/her family and other parties involved in a case, and organizing services. The social work managers occasionally took part in client work. They gave consultancy to their subordinates, carried out some official decisions and reacted to clients’ contacts.

CIS had an integral role in managing clients’ cases. Each child has their own file in CIS (Fig. 1). The current CIS was introduced to our research sites A and C in 2004 and in site B in 2006. Previously, another system was in use. In the implementation stage, only the basic information of active clients was manually entered to current CIS. Because of the lack of automatic conversion, narrative case reports were, however, left in former CIS. Each research site tailored CIS to match the local needs.

CIS has separate modules for documentation, structural family information, decision making, scheduling and statistical presentation of work activities. In the documentation module, social workers keep up the chronological case report and write a structural client plan. Family module depicts the people living in the same household. Decision making module is used, for example, when services are granted for a client. Scheduling module presents the forthcoming appointments. Work done module is used to numerically describe the past actions. Each module can be purchased separately. Therefore, there is a slight variation in what modules are available across research sites.

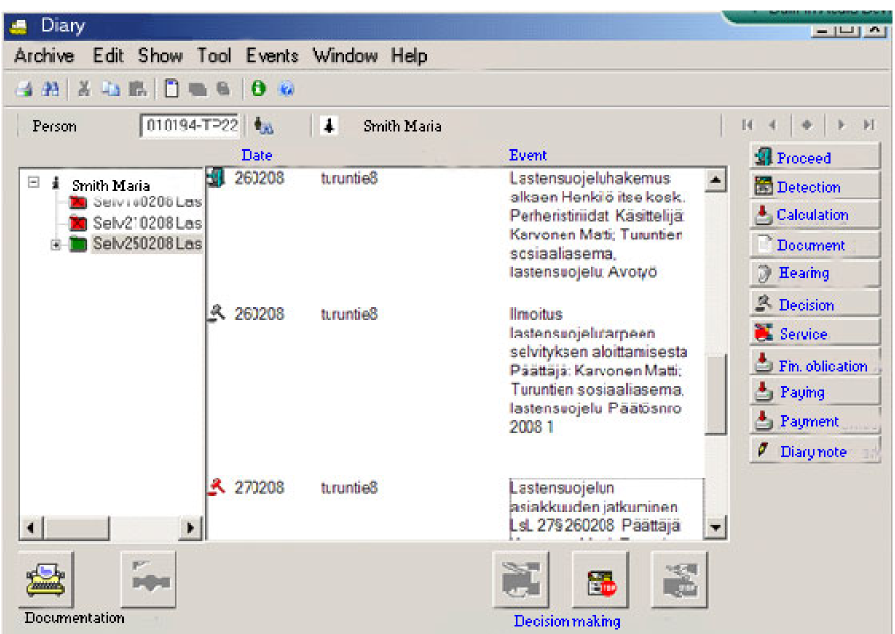


Fig. 1. The diary view of CIS - translated in English (fictional client)

The so called diary view (Fig. 1) of a client serves as an access point to different modules. There is no possibility to navigate directly from one module to another but a loop through diary view is needed. The navigation buttons to the modules can be seen on the right side of the screen capture. The diary view also shows different files (on the left) and offers a modest glance at recent decisions of a case (on the middle of the screen). CIS offers limited search tools for the social workers. It is possible to search clients by their name or ID. However, there is no search option within a case report.

4.3 Data Collection

The study employed semi-structured interviews and field observations. The interviews were carried out first. Willingness to participate in observations was asked during the interviews and the actual observations were realized later. In city A and B the majority of social workers were interviewed whereas in city C about one fourth of the workers participated. The interviewed social workers had varying work histories in child protection services. There were novices, those who had a couple of years work history and a few having a long term career. The observed social workers had a heterogeneous history of using CIS. Some had experience only from this particular CIS and some had used various types of systems previously.

Before the actual interviews, a pilot interview was conducted to check the functionality of the questions and time required for an interview. Our data consists of 33 semi-structured interviews with social workers (23), social work managers (7) and system administrators (3). The interviews took place at the offices of the social workers, with only one exception at the university. The core themes of the interviews were work related to child protection, documentation and use of CIS as part of work. In sixteen interviews, social workers described an individual client's case by drawing a timeline and network map about persons involved in a case. The cases had a role of a critical incidence [29] allowing more focused and concrete questions. The total interview data is about 62 hours. One interview is about two hours in length. The interviews were tape-recorded and fully transcribed.

Twelve field observation sessions were realized in real-life settings. Eleven social workers were observed, one social worker twice. The social workers have a practice to book up time to do documentation and other work with CIS. The social workers were asked if it was possible to join those sessions and follow their work. The social workers worked normally in these observation sessions. The only exception was that based on their own choice they explained and reflected about their actions concurrently every now and then. This way, a verbal protocol [11] was produced. Observation data consists of hand written field notes from all sessions and the tape-recorded commentary by the social workers from ten sessions. In two sessions it was not appropriate to do recording. Field notes were written up shortly after the observation session. Later on, the commentary of the social workers was transcribed word for word. The observation data is about 40 hours. The sessions varied from about two hours to the whole working day.

4.4 Data Analysis

Firstly, the interviews were read through several times to become familiar with the data. After that, appropriate parts of the data were collected under four wide themes according to research questions. The data extractions were thoroughly scrutinized and summarized with few words in the tables. This enabled us to analyze the dimensions of phenomena in a more itemized way.

Each critical incidence - client case used as an example in the interviews - was pulled together so that it formed a consistent account. This meant that all events were listed chronologically. Also institutional actors and people related to clients' private life were listed. The use of information and CIS was identified and itemized in the critical incidents.

From the interviews, those of the front line social workers' and social work managers' were used as a primary source. The interviews of systems managers offered background information to understand the properties of CIS.

The notes from the observations and social workers' commentary during the sessions were combined. Thereafter, the data was organized according to functional entities. A functional entity is a single task in CIS consisting of a sequence of actions occurring in order to reach a certain goal. The functional entities include activities like writing a case report and making different types of decisions. These functional entities were studied more carefully from two aspects: how social workers keep up the temporal trajectory and how they look for a temporal overview of a case.

The different datasets shed a different kind of light on the research questions. The interviews offered an overall picture of significant information and a role of CIS as part of work. The critical incidences within the interviews gave a shape to the clients' trajectories. Observations and social workers' verbal protocol concretized the human-computer interaction.

We take two standpoints to temporality when we present the results. Firstly, we describe the temporal trajectory in general. Secondly, we link temporal aspects (episode, interval, eon) [26] to information presented in the trajectory.

5 Results

First, we describe information that the social workers consider as important for understanding a client's trajectory. Next, we analyze how this information is presented in CIS of our research sites, and the capability of CIS to give a temporal overview of a client case. Lastly, we briefly discuss social workers' strategies of producing and using the temporal trajectory in CIS as part of their daily work.

5.1 Information Strands in a Temporal Trajectory of a Case

Social workers clearly highlighted the importance of seeing and interpreting a client's situation as a temporal continuum. This meant a chronological and consistent narrative that could depict what had been the starting point and what had happened during a child protection process. The need to observe changes in the process was expressed by a social worker: "*What has changed, if there is something new. I aim to depict the trajectory.*" (O 1)

However, it turned out to be more difficult for the social workers to define what kind of information belongs to the temporal trajectory. It is not about working around a single issue but rather handling a case from a much broader perspective. As expressed by one of them: *"We manage a person's life in its entirety so what would not be meaningful."* (3) A full trajectory is constructed from several information elements. Trajectory management is reminiscent of piecing a jigsaw puzzle together. The core of the problem is: *"Also those small pieces of information are important so it is not enough to put down summarily those big outlines."* (1)

The client's records serve as a temporal boundary object. Information is recorded in a specific moment and it might be used months, even years later. Often, however it is unclear what information is needed in the future. Besides the wide working orientation, situational and temporal factors also seem to somewhat explain what is considered as relevant information.

Three major information strands of the temporal trajectory were identified. They were strands of concern, a client's private life and institutional actions. They were intertwined. To understand a client case as a whole requires seeing the strands together.

Concern Strand. To the social workers, the concern strand explained the varied reasons for the beginning and continuation of customership in child protection services. The social workers appreciated concrete expressions of concerns. The essential was to find out: what is the concern about? The list of major reasons for concern was relatively limited. Moreover, often it is not possible to articulate well what it is all about.

It is common that the concern strand has several actors expressing their point of view about a client's situation. In some cases, opinions about concern can be contradictory, and also contradict the social worker's opinion.

The temporal strand of concern is not necessarily linear, stretching over the past and the future, as the examples told by social workers demonstrated. The arc can be cyclic when there are passive and active stages of concern. The concern can fluctuate when a level of concern takes turns. The course of concern might be steady or cumulative with increasing issues of concern coming out. The concern strand can consist of several of these features.

Concern is hard to fit into any of the temporal aspects (episode, interval, eon). It is rather an all over state continuing across the temporal trajectory. In that sense, it belongs to the temporal aspect called eon. However, the level of concern can have peaks or periods of good or bad times. In those situations, the episode and interval aspects of time are applicable.

Private Life Strand. The essential information in the private life strand concerns the whole family, not only the child. We recognized three major information themes that social workers talked about: major events, everyday life at home, and family as a network. Noteworthy is that information is often needed about the past events and life situation from the time before the child protection. After all, the private strand describes "the history of family".

All the aspects of time match up with the private strand: episodes for events, interval for certain periods in family life and eon to depict continuity in everyday life.

Institutional Strand. The reason and level of concern determine what kinds of services are organized for a child and family, and what institutional actions are taken. The institutional strand describes the work done in the child protection. The social workers seemed to need a long term view of services: what had been in use but terminated, how did services get started, what currently is going on, how long were the service periods, and what are the plans for the future. The evaluative information about the impact of services and family attitudes towards services were needed as well. The analysis of the network maps of the critical incidents revealed that social workers co-operate with a multitude of people. There were co-workers within their own social service organization (emergency department, adult services, welfare for intoxicants) and outside of it. All of them are potential information producers.

From the temporal aspects, episode and interval seem to felicitously depict the institutional work.

To conclude this chapter, we use data extracts to demonstrate the temporality and a contingencies, in the work.

”Q: What is the essential information in child protection, what should be available from each client?
A: Well, why the customership exists and what is the goal and what is the plan to reach the goal.
Q: Hmm.
A: Often we are reproached that we don’t have any plans in our work. We are making plans today and in the next day our plans might be vitiated because of a single phone call.” (6)

We can summarize the core findings of information strands in a temporal trajectory to the following: reason and level of concern; services and actions child protection has taken plus their impact; major events, everyday life, and a child’s network are core information needed in the overview to support social workers in their work. The intertwined strands have different temporal aspects and lengths, the institutional trajectory maybe being the simplest one to represent. In the next chapter, we take a closer look at how CIS is capable of presenting the three identified strands as a trajectory.

5.2 Presentation of Temporal Trajectory in CIS

We mirror information presentation in CIS through a specific case of Maria. The temporal trajectory (in italic in the next paragraph) is the idealization of several critical incidences presented in the interviews. The strands of the temporal trajectory - concern, private life and institutional – are narrated side by side.

Maria’s customership was opened at the beginning of 2005. The school contacted child protection. Maria attended school irregularly and looked unbalanced. Her parents were divorced. She lived with her mother and two younger siblings. The social worker called the teacher and the mother. There were meetings with Maria and the mother at the office and a visit to home. The social worker noticed that the mother seemed to have health problems, balanced with short term jobs with no time and energy for children. The social worker organized a support family (obtained after six months waiting) and financial support for Maria’s hobbies. The mother’s partner moved in and out, the family moved within the city, Maria participated in a summer camp. Everyday life rolled on. In 2006, a new social worker took care of Maria’s case. Meetings were irregular. At the beginning of 2007, the neighbours made two child welfare notifications: children had been alone at home. The third social worker *visited the family. She could not figure out the situation, just made observations about home and the mother's tiredness. To strengthen the support, a family worker started to visit the family. However, the work was soon discontinued because the mother wanted it to be. In 2008, customerhsip was on hold. Then the social worker got a very serious child welfare notification by anonymous phone call – it was time to react.*

Next we take a look at how information about Maria's trajectory is distributed, and can be found within CIS: in modules of documentation (includes e. g. case report, client plan), of family, of decision making, of scheduling and of work done.

Concern Strand. There are four child welfare notifications. The first one is from the school, two later ones from the neighbours and the last anonymous one. The other form of concern is that of the social workers: the observations about the mother's tiredness in the first place. Later it is only a hunch when the social worker is not able to clarify a crux of the matter. The notifications have their own registry in CIS, where they are recorded. This information is brief and structured: a date, the name of informant and the content of contact. In most cases, the social workers also write as a link a short entry about the notifications to the case report. The social workers' own concern and assessment of its level is embedded in the case report. The worry expressed by different actors during discussions is also in the case report.

Private Life Strand. There are few events listed in Maria's and her family's life: a summer camp, changes in the mother's relationship and a change of residence. We can also find references to the time before customership. In the past the parents have divorced. There is also information about everyday life. These are reported on the case report. In addition, the official changes in the household composition (move of the mother's partner) and a move during the customership are depicted in structured form in the family module.

Institutional Strand. During the course of the trajectory there were meetings, phone calls and home visits. Besides events, the social workers made four formal decisions. The first one is about to start the customership and three latter ones about the services: financial support, a support family and family work. The staff turnover is also a part of institutional strand: three different social workers took care of Maria's case. The institutional work is fully reported on the case report in the documentation module. The entry of a short event starts with the date, social worker's name, the name of event and the listing of participants. Thereafter, the content of the event and discussions are described. It is a practice to mention the decisions made in the case report. The institutional work is monitored in three other modules. In the decision module are the structured decisions including written justification. The work done module indicates the already realized actions (phone calls and meetings etc.). It briefly depicts the type of action, the number of participants and time in minutes spent with a work task. The scheduling module orientates to the future. It indicates the forthcoming meetings and home visits. These modules, however, give only a limited view of the work.

Obviously, the case report on the documentation module offers the most comprehensive account of Maria's case. It binds all three information strands together. Its temporal perspective covers the time before child protection, current time and the future. The information in the other modules is mostly used for administrational and accountability purposes. Their timeframe covers only the current time in the child protection. The information presentation is multifold since the same information can be found in different modules expressed in a different form. The information strands, modules containing information about them, and their temporal aspects in CIS are summarized in Table 1.

Table 1. Information strands, modules and temporal aspects in CIS

Strand	Module	Temporality
Concern	Documentation Notification registry	episode, interval, eon – past, current, future episode - current
Private life	Documentation Family module	episode, interval, eon – past, current, future episode - current
Institutional	Documentation Decision module Work done module Scheduling	episode, interval, eon – past, current, future episode, interval - current episode - current episode –current (orientation to future)

According to the social workers the fastest way to get an overview about a client's trajectory was to read through a client plan. It summarizes the history of the family and clarifies the client process and goals set. In the client plan information is ready selected and analyzed. However, not all clients had one. In those cases, the social workers were dependent on a case report if they wanted to figure out a child's history as a whole.

Overall, the significant issue was the difficulty of identifying the thread running through the case report: finding the course of action. This was particularly the case in active client processes when reports could be "thick like a brick". The social workers noted that the concern, major events, goals and their realization did not come up clearly enough. The case report contained those small pieces of information recorded just in case. The other reason is that the history is made up of different entries as life unfolds.

CIS could give only superficial support depicting an overview of the case. The diary view visualizes only the different files and major decisions. The following data extract exemplifies not only how information comes up in the case report but also the social worker's fast interpretive skills. She discusses information through institutional strand: numerous entries are a message from active working that likely indicates the seriousness of a case. By looking at the time frame she analyzes the shape and stages of a trajectory. The work of the emergency department is a message about possible crisis and contingencies.

*"You can at a glance see **how many entries** there are [in a case report]. Sometimes it scrolls a long time to get open. Then you know that there is plenty of writing and a long customerhip. You can take a look at **what time frame those entries have been made**…You can see from the initials if it has been the **emergency department** working over the case. You can make those short observations."* (4)

Client information was spread over several modules in CIS. Information of institutional strand was in four modules, and concern and private life strands in two modules. In the documentation module, case report compiled information of all strands. However, it was difficult to get an overview about a client's case based on it.

5.3 Interacting with CIS: Managing and Using a Temporal Trajectory

Social workers had different strategies of managing and using trajectories when they interacted with CIS. Since the major source of information was the case report, the majority of the management work was done there. The social workers' goals were first of all, to keep the trajectory as a whole in the case report by referring to possible information in other modules. Secondly, there were different types of attempts to point at the essential information in the case report. To gain a temporal overview of a case, social workers approached the case report either by skimming text through or by reading intensively.

***Managing Trajectory*.** We identified three tactics to keep up the temporal continuum in the case report. *Firstly*, the social workers created *links to information* outside of the case report by writing short references. They indicated whether there was more information in another document (e. g. in the client plan) within the documentation module or in another module. In the latter case, references were mostly made to decisions that existed in the decision module. The social workers also created traces to hint that more information could be found from the family members', siblings' or parents' files in CIS. In the client cases that had started before the implementation of CIS, there were occasionally references to old CIS. Often, the social workers made a note if there was more information in paper files, too. *Secondly*, the social workers *wrote*, although very rarely, *summaries.* The summaries related to the past, a time before a child became a customer in this particular organization. This was the case, for instance, when a child had a history in another municipal's child protection service or there were previous remarks concerning the family. Occasionally, summaries were made about paper reports. *Thirdly,* in some situations the social workers kept the case report consistent by keeping up a *dialogue between* past and current *entries* and between the entries of co-workers. The current entries were reactions to the past ones. The linking strategy (manually added notes to the case report) was relatively simple and more technical by nature, instead summary writing is rather a writing practice.

Indicating the essential information was realized by adding titles to structure the text, using bold fonts or making listings to the end of entries. The bold fonts were used, for example, to highlight turning points or serious events. Serious issues that were raised for discussion for the first time could be entered in bold text. The social workers used listings in the entries, mostly to summarize agreements and actions to be taken in the future. These tactics aimed at offering handles for fast access to information.

Using Trajectory. We identified two tactics, holistic check up and intensive reading, as the social workers' ways of familiarizing themselves with the whole trajectory in CIS. The holistic check up means skimming through the text in the case report and trying to locate the essential parts of it. In the intensive reading the social workers carefully go through the entries. In some cases, they print out the case report and rather read this printed document.

The time available to read the entries, the familiarity of a client, a work task or a work situation at hand were some of the reasons that dictated the ways of reading. The social workers skimmed through the past entries before the meetings, when they were on the phone with a client or when they evaluated the continuity of a client's case quickly. The previous information had a special significance to the social workers who took over an old case and tried to figure out the past. The taken into care decision or making a first client plan for a long-term customership required intensive reading.

We illustrate these two information use tactics by presenting two functional entities from the observation sessions. On the tables, the left column describes the major moves made in CIS. The right column includes the essentials of the verbal protocol. The fist example depicts the holistic check up and the second one intensive reading.

In the first example (Table 2), the social worker goes through a client's list in a semi-annual customers' 'check-up day'. She evaluates whether the customership needs to be sustained or be terminated. She navigates from the child's file to the mother's file, returns to the child's file to close down the case and finally closes down the sibling's case too. She looks at information from the temporal perspective. Firstly, she looked at the starting time, and then stages in the client process. She notices that the situation is passive. She uses all information strands of the trajectory. She makes observations about the services that had been in use (institutional strand), reflects the concern and finally seeks consolidation from the mother's income support file (private life strand). She remembers the mother from engagements at the office. The information in CIS combined with her own experiences support decision making. She decides to close the case and the sibling's case as well. The task does not require much time.

Table 2. Holistic check up – evaluating the customership

Moves in CIS	Verbal protocol
1. skims the case report of the child **2.** skims the case report of the mother **3**. checks the income support calculation of the mother **4**. returns to the child's file and terminates the customerhip **5.** moves to the sibling's file and terminates the customerhip	*"This has happened in 2004. This case is still open. There has been family work but terminated. Well, in 2006 – now I got hang on this. They have been in a summer camp. I remember that the mother has been here for income issues.* ***The past action was two years ago. I don't have any child protection work here****… I take a look by mother's name…* ***There is neither work going on nor is there concern.*** *"*

Our second example (Table 3) draws from the situation in which the social worker prepares a tentative client plan to be discussed in a meeting. The child has the history of several years in the child protection. The social worker has been in charge of the child's matters for about a year. According to her, the customerhip is activated because of a crisis. For the plan, she must piece together a compiled package about concerns, private life and institutional actions. She starts by glimpsing a case report and correcting some spelling mistakes in it. Thereafter, she prints out the case report's entries from the past year. Then she moves to the actual task at hand: preparing the client plan. She reads the case report both from the CIS screen and from print. She tries to figure out the course of action. She is lost in the text: she needs to locate information piece by piece. As she says, information is fragmented in the case report. She makes notions about services (institutional strand), child welfare notifications (concern strand) and family situation (private life strand). She finds out that the situations of the same kind reoccur and continue in the family's life. That is an important piece of information. She copies and pastes between the case report and the client plan. She spends most of the afternoon familiarizing herself with the client's case. To boot, there would be even more information in old CIS and paper files.

Table 3. Intensive reading – preparing a client plan

Moves in CIS	Verbal protocol
1. skims the case report of the child **2.** checks assistance (how to check spelling in CIS) from a colleague **3.** prints out the case report from the past year **4.** works with a client plan **5.** reads the case report carefully **6.** copy-pastes between these two documents **7.** prints out the half-done plan	*"I need to ask if somebody can take a look at information* ***in old CIS.*** *You just have to search for the information. If a case report is all you got you are totally lost.* ***This is quite a patchwork.*** *If only there could have been a way to pick up [the services] somehow. By now, I have spent almost 1 ½ hours on this.* *I could take a look at the* ***paper files*** *too. What had happened in the child's life, what has it been, the background information is laborious to find.* ***The information is fragmented there****."*

The two previous examples depict the role on information and CIS as part of two different types of work task. The social workers interact with information and CIS alike. In both cases, the social workers search information from the perspectives of three information strands. They also make notions about the past and the current situation of a case when trying to assess the possible future of a child. The tasks require both the identification of essential information and navigation between different modules in CIS. Apparently, CIS could offer more support to realize the tasks. However, it is a matter of the social worker's professionalism, too. As expressed by the social worker when talking about whether there was enough information for preparing a taken into care decision: "*Yes there was, but you have to see the wholeness. And [understand] what to do with that information.*" (7)

6 Discussion and Concluding Remarks

Summarizing the Findings. In child protection, there is a need for case overviews presented by CIS [8]. The overview would support the social workers' work that is framed by limited time resources, heavy caseloads and situations when immediate reactions are needed. We approached the construction of the overview from three perspectives. Firstly, we identified the core information strands belonging to the case overview. Secondly, we analyzed how these information strands were presented in one particular CIS. Thirdly, we observed how the social workers managed and used these strands in CIS.

In the previous studies, CIS has been considered as a threat to presenting holistic information in social work [7], [16], [33]. Holistic information has been taken for granted with no exact explanation of what it is and how it should be presented in CIS. To design an overview, however, its core information elements should be defined. We identified three interconnected information strands in a client's trajectory: of concern, of private life and of institutional actions. The social workers preferred to see the strands from a temporal perspective. They needed information about the past, the present and the planned future of a child. The temporal continuum also describes change, continuity and permanency. Visibility of these three factors is central when trying to understand the complexity of case work [19].

The information elements of the concern and private life strands included three temporal aspects: episode, interval and not so clearly time sliced aspect, eon. Instead, information elements in the institutional strand included the aspects of episode and interval. Consequently, the information in the institutional strand is easier to describe with exact time stamps than in the two other strands. Time has been modeled according to instant (aka episode) and interval in information system design [4]. Our findings indicate that information can be temporally depicted similarly in CIS in social work.

Information was distributed across separate modules in the CIS analyzed. However, the case report had a central role to keep three information strands together. The report is a sequential account [31] but still, the entries can refer to the past, current and future issues in a child's life. Thus, it is not only the CIS fragmenting information as has often argued [7], [16], [33] but as well the nature of work that generates information as time sliced pieces. The social workers can only write the child's story as it unfolds. Information is obtained piecemeal during a client process.

The social workers managed a client's trajectory within a case report in two ways. Firstly, they created references to information existing in other modules or paper files. They did this to ensure a temporal continuum of a case at least in one place. Secondly, the social workers occasionally highlighted the essential information to create help in identifying the essentials. They used their own tactics to patch up the missing properties of CIS.

The social workers had to piece together an overview of a case in their minds. They either tried to scan quickly through a case report (holistic check up) or if necessary read through a case report carefully (intensive reading). The previous studies on medical [15] and social work [12] context identified skimming as one approach to documents. Obviously, exhaustive information is not always needed. Even a hint, e.g. bolding, can fill the information gap and guide in work task performance.

Implication for Design. There has been growing interest toward temporal representation of information. Proposals have been made to visualize information in timelines, e. g. with patient data [6], [30], personal histories (LifeLines) [18] and newspaper news [10]. Our findings are in line with these proposals. The social workers themselves directly referred to diagrams as alternative ways to depict the core history of a child. Also, assuring a temporal continuum of different documents is a way to keep a child's case history coherent.

Three information strands found could be depicted as parallel timelines. The construction of timelines requires collecting information in different CIS modules together. The timelines could have 'pointers' with snippets to actual documentation. The social workers could choose whether to draw on a summary or navigate to the actual documentation. These alternatives would support two reading tactics: skimming and intensive reading. An opportunity to mark possible contingencies [27] or turning points are similar to the social workers' currently used tactic to highlight issues in a case report.

The child welfare notifications are concrete expressions of concern to be compiled to the concern timeline. The challenging part, however, is to depict the social workers' own assessment or hunch of concern. As carefully proposed in a couple of the social workers' interviews, the concern could be 'measured' and indicated with colour codes: red for high and green for low concern level. It is a double-edged sword though. The coding might be supportive but it also adds to the work of social workers.

Private life constitutes events that are, at least partially, possible to collect to the timeline. Family is a central context of a child's private life. It includes siblings, parents, stepparents, grandparents etc. Depicting a family as a social network graph might ease the understanding, in some cases, of complex family structures. In many cases, the social workers navigate between family members' files. The network graph could offer access points between the files.

The timeline of institutional strand could depict given service periods and actions taken by social workers and other workers. This timeline would support seeing a cumulative arc of action on a trajectory [27] by showing all the provided services.

Creating a continuum of documents requires collecting entries from different modules and scanned paper statements to a 'document timeline'. Scanning possibility exists already now in CIS. Currently, the social workers sustain the continuum of documents in a case report by their linking practices.

System design should be done keeping in mind that social workers need a tool that is simple and easy enough to use [8]. The possible new features in CIS should not produce an extra burden for social workers.

Implication for Research. The construction of a timeline can be based on existing temporal markers in CIS. Taking a look inside at different documents, especially a case report, is essential. Thus, further analysis of temporal markers is needed.

The social workers take a lot of effort to produce information about a child's case in CIS. Currently, CIS does not offer a coherent overview of a case that is available at a glance. It is possible to design a holistic overview of a case to match the social workers' needs and to support social workers' work.

Acknowledgments. This study is financed by the Academy of Finland grant 133053. We thank the employees at our research sites for participation, and Tarja Pösö, Aino Ritala-Koskinen and Tarja Vierula in our CHILDINFO research group for helpful comments.

References

1. Aas, K.F.: From Narrative to Database: Technological Change and Penal Culture. Punishm. Soc. 6, 379–393 (2004)
2. Ash, J.S., Berg, M., Coiera, E.: Some Unintended Consequences of Information Technology in Health Care: The Nature of Patient Care Information System-Related Errors. J. Am. Med. Inform. Assoc. 11, 104–112 (2004)
3. Bell, M., Shaw, I.: Integrated Children's System Evaluation: Summary of Key Findings. DCSF-RBX-02-08 (2008)
4. Combi, C., Keravnou-Papiliou, E., Shaher, Y.: Temporal information systems in medicine. Springer, New York (2010)
5. Compton, B., Galaway, B., Cournoyer, B.: Social work processes, 7th edn. Thomson Brooks/Cole, Belmont (2005)
6. Drohan, B., Grinstein, G., Hughes, K.: Oncology Lifeline - A Timeline Tool for the Interdisciplinary Management of Breast Cancer Patients in a Surgical Clinic. In: Information Visualisation (IV), 14th International Conference, pp. 233–238 (2010)
7. Hall, C., Parton, N., Peckover, S., et al.: Child-Centric Information and Communication Technology (ICT) and the Fragmentation of Child Welfare Practice in England. J. Soc. Policy 39, 393–413 (2010)
8. Huuskonen, S., Vakkari, P.: Client Information System as an Everyday Information Tool in Child Protection Work. In: Proceeding of the Third Symposium on Information Interaction in Context, pp. 3–12 (2010)
9. Lie, M.: Technology and Gender Versus Technology and Work: Social Work and Computers. Acta Sociol. 40, 123–141 (1997)
10. Matthews, M., Tolchinsky, P., Blanco, R., et al.: Searching through Time in the New York Times. In: Proceedings of the Fourth Workshop on Human-Computer Interaction and Information Retrieval, pp. 41–44 (2010)
11. Miller, R.L., Brewer, J.D.: Verbal protocol analysis. In: Miller, R.L., Brewer, J.D. (eds.) The A-Z of Social Research. Sage, Thousand Oaks (2003), `http://srmo-dev.ifactory.com/view/the-a-z-of-social-research/n117.xml`
12. Monnickendam, M., Yaniv, H., Geva, N.: Practitioners and the Case Record - Patterns of use. Adm. Soc. Work 18, 73–87 (1994)
13. Munkvold, G., Ellingsen, G.: Common Information Spaces Along the Illness Trajectories of Chronic Patients. In: Bannon, L.J., et al. (eds.) ECSCW 2007: Proceedings of the 10th European Conference on Computer Supported Cooperative Work, pp. 291–310. Springer, London (2007)
14. Murphy, J.W., Pardeck, J.T.: Computerization and the Dehumanization of Social Services. Adm. Soc. Work 16, 61–72 (1992)

15. Nygren, E., Henriksson, P.: Reading the Medical Record. I. Analysis of Physicians' Ways of Reading the Medical Record. Comput. Methods Programs Biomed. 39, 1–12 (1992)
16. Parton, N.: Challenges to Practice and Knowledge in Child Welfare Social Work: From the 'social' to the 'informational'? Child. Youth Serv. Rev. 31, 715–721 (2009)
17. Pithouse, A., Hall, C., Peckover, S., et al.: A Tale of Two CAFs: the Impact of the Electronic Common Assessment Framework. Br. J. Soc. Work 39, 599–612 (2009)
18. Plaisant, C., Mushlin, R.: An Information Architecture to Support the Visualization of Personal Histories. Inf. Process. Manage. 34, 581–597 (1998)
19. Pösö, T.: Revisiting Residential Care. Qual. Soc. Work 9, 27–42 (2010)
20. Reddy, M., Dourish, P., Pratt, W.: Temporality in Medical Work: Time also Matters. Comput. Supported Coop. Work 15, 29–53 (2006)
21. Reddy, M., Dourish, P., Pratt, W.: Coordinating Heterogeneous Work: Information and Representation in Medical Care. In: Prinz, W., et al. (eds.) ECSCW 2001: Proceedings of the Seventh European Conference on Computer Supported Cooperative Work, pp. 239–258. Kluwer, Dordrecht (2001)
22. Sapey, B.: Social Work Tomorrow: Towards a Critical Understanding of Technology in Social Work. Br. J. Soc. Work 27, 803–814 (1997)
23. Savaya, R., Spiro, S.E., Waysman, M., et al.: Issues in the Development of a Computerized Clinical Information System for a Network of Juvenile Homes. Adm. Soc. Work 28, 63–79 (2004)
24. Savolainen, R.: Time as a Context of Information Seeking. Libr. Infor. Sci. Res. 28, 110–127 (2006)
25. Shaw, I., Bell, M., Sinclair, I., et al.: An Exemplary Scheme? an Evaluation of the Integrated Children's System. Br. J. Soc. Work 39, 613–626 (2009)
26. Sonnenwald, D.H., Iivonen, M.: An Integrated Human Information Behavior Research Framework for Information Studies. Libr. Infor. Sci. Res. 21, 429–457 (1999)
27. Strauss, A.L.: Continual permutations of action. AldineTransaction, New Brunswick (2008)
28. Strauss, A.: Social organization of medical work. University of Chicago Press, Chicago (1985)
29. Urquhart, C., Light, A., Thomas, R., et al.: Critical Incident Technique and Explicitation Interviewing in Studies of Information Behavior. Libr. Infor. Sci. Res. 25, 63–88 (2003)
30. Wang, T.D., Wongsuphasawat, K., Plaisant, C., et al.: Visual Information Seeking in Multiple Electronic Health Records: Design Recommendations and a Process Model. In: Proceedings of the First ACM International Health Informatics Symposium, pp. 46–55 (2010)
31. White, S.: Time, Temporality and Child Welfare: Notes on the Materiality and Malleability of Time(s). Time Soc. 7, 55–74 (1998)
32. White, S., Hall, C., Peckover, S.: The Descriptive Tyranny of the Common Assessment Framework: Technologies of Categorization and Professional Practice in Child Welfare. Br. J. Soc. Work 39, 1197–1217 (2009)
33. White, S., Wastell, D., Broadhurst, K., et al.: When Policy o'erleaps itself: The 'tragic Tale' of the Integrated Children's System. Crit. Soc. Policy 30, 405–429 (2010)

Unsupervised Parameter Selection for Gesture Recognition with Vector Quantization and Hidden Markov Models

Przemysław Głomb, Michał Romaszewski,
Arkadiusz Sochan, and Sebastian Opozda

Institute of Theoretical and Applied Informatics of Polish Academy of Sciences,
Bałtycka 5, 44-100 Gliwice, Poland
przemg@iitis.pl

Abstract. This article presents an investigation of a heuristic approach for unsupervised parameter selection for gesture recognition system based on Vector Quantization (VQ) and Hidden Markov Model (HMM). The two stage algorithm which uses histograms of distance measurements is proposed and tested on a database of natural gestures recorded with motion capture glove. Presented method allows unsupervised estimation of parameters of a recognition system, given example gesture recordings, with savings in computation time and improved performance in comparison to exhaustive parameter search.

Keywords: Gesture recognition, Vector Quantization, Hidden Markov Model, automatic parameter selection.

1 Introduction

Human-computer interfaces (HCI) using natural human gestures are one of the promising new approaches in interaction design. The important component of a gesture interface is a recognition system for time series measurements of data (motion capture readings or image derived features). For this task, one popular and reported as effective approach is the combination of Vector Quantization (VQ) and Hidden Markov Model (HMM) [1,2]. While effective, this method needs to be provided with several parameters, notably number of VQ symbols, number of HMM states, and initial choice for HMM matrices. Those parameters are often chosen ad hoc or by trial and error, with random selection of initial conditions. Such approach is time consuming and can lead to suboptimal performance [3,4]. On the other hand, the performance of input recognition largely determines the usability of a HCI system.

Our approach and main contribution of this work is the investigation of a two step, heuristic algorithm for VQ+HMM parameter selection: first, computing the number of VQ symbols by considering class separation, followed by iterative hierarchical clustering of subsequences for finding initial HMM matrices. The standard model reestimation (Baum-Welch algorithm) and selection (Akaike/Bayesian information criterion or others) can be applied afterwards for completing the parameter selection. This algorithm allows for deterministic and unsupervised estimation of parameters of

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 170–177, 2011.

a recognition system, given example set of gesture recordings, with savings in computation time and improved performance in comparison to exhaustive or random parameter search.

1.1 Related Work

The choice of parameters is an integral element of a development of a VQ+HMM system, and as such has been a subject of active research. In [11], one of the most popular Vector Quantization algorithms was introduced, named after authors' initials as LBG. By an iterative process of splitting and optimizing cluster centers, it finds optimal VQ dictionary vectors with respect to the distortion measure. This approach produces good results for coding applications, but does not consider the discriminative power of the sequences of symbols for the classification task. The latter problem can be related to estimation of class separation and within/between class variability as considered in Linear Discriminant Analysis (LDA). The problem of parameterization of VQ for optimal class separation has been considered in [12] using Dynamic Time Warping (DTW) and Learned Vector Quantization with the objective of speech recognition with Self Organizing Maps. The similarity of symbol sequences with application for recognition has also been the subject of [13], where DTW was applied to sequences after VQ. The problem of optimization of VQ centroids has been addressed in [14], with application of Bhattacharyya distance.

Our approach builds on above works, focusing on maximization of the Bhattacharyya measure of within-class and between-class distance distributions, obtained using Levenshtein symbol distance, a technique conceptually similar to DTW.

The other problem with design of VQ+HMM recognizer, apart from the above, is the parameterization of a HMM. [1] note the importance of good initial HMM values, as the reestimation algorithm (Baum-Welch) finds local, not global maximum of likelihood function; a procedure is thus proposed for improving the initial values using K-means clustering of data vectors. The clustering can be also used to identify the number of HMMs needed for representation, as in [8] where DTW and hierarchical clustering is used to obtain a set of HMMs describing a set of human motion trajectories. In [2] authors optimize the performance of a HMM by merging states based on Kullback-Leibler similarity measure of symbol emission matrices, however, this is done after training, with the chief objective of reducing number of states of a HMM constructed as a sum of several others. In [5], several model selection strategies are investigated, including sequential state pruning strategy, where least probable states are iteratively deleted from the model. While effective, deletion could possibly lead to a loss of information in the final model. [6] perform an in-depth analysis and argue for using cross-validated likelihood to assess the number of states.

Our approach considers estimation of initial HMM parameters, which are to be improved with Baum-Welch optimization. We split the sequences into fragments and successively merge them with agglomerative hierarchical clustering based on the Bhattacharyya distance of distributions of symbol occurrences. At each moment, the clustered fragments can be translated into HMM initial parameters.

This article is organized as follows: section two contains the method description, section three experimental results with motion capture database, last section presents conclusions.

2 Method

We focus on the problem where the user interface (UI) designer has the following:

1. An acquisition system that captures human motion as a time sequence of vectors $X=(\boldsymbol{x}_t)$, where each $x_t \in \mathbb{R}^l$ represents the data from l sensors (finger bend, hand rotation, and similar when using motion capture device) or features (such as trajectory features, shape moments when working with images of motions).
2. A set of c gestures (defined motions) selected to be used in the interface.

The input sequence X is classified into one of the c gestures using VQ+HMM framework as follows. First, a quantization function

$$Q: \mathbb{R}^l \to \mathbb{Z}_m \quad x_t \to s_t \tag{1}$$

is used, transforming a sequence of vectors X into a sequence of discrete symbols $S=(s_t)$. Then, a set of HMM models corresponding to the gestures $\lambda_1,\ldots,\lambda_c$ is used to compute the estimate of the sequence being generated with each model. Either the forward or Vitterbi algorithm can be used for this purpose [1]. By comparing the result probabilities the class of the input can be derived.

Implementation of this approach requires definition of the VQ and HMM stages. For VQ a quantization algorithm must be selected, for example the Linde-Buzo-Gray (LBG) method [11]. Given a set of reference data, the algorithm can calculate the m reference ("dictionary") vectors. The number of VQ symbols m (or related parameter, like target distortion ε), however, has to be explicitly provided. For HMM, the defining stochastic matrices (starting state probability vector Π, state to state transmission probabilities A and symbol emission probabilities B) must be given. The HMM matrices are commonly estimated with Baum-Welch method [1], but this requires definition of their initial values and implicitly providing the number of states n.

2.1 Choosing the Number of Vector Quantization Symbols

The objective of this stage is to derive the number of VQ symbols m. The proposed method follows from the observation that for any two recorded symbol sequences S_1 and S_2 we can use approximate string matching to compute the distance $d_S(S_1,S_2)$ between them. The use of approximate methods is necessary, as subsequent recordings of the same gesture usually differ in both length and content of the symbol sequence, resulting from variations of movement (speed, hand configuration and so on). The nature of those differences can be formalized as symbol addition, deletion and replacement, thus Levenshtein distance is particularly fitting for this purpose.

Given a set of c gesture classes with k realizations of each class we can compare them both within their classes (by computing $d_S(S_1,S_2)$ where S_1 and S_2 belong to the same class) and between them (where each of S_1 and S_2 belong to a different class).

We can compute the histograms h_W of within-class d_S and h_B of between-class d_S. After normalizing the histograms, we can compute Bhattacharyya distance as

$$d_B(h_B, h_W) = -\ln \sum_{i \in I} \sqrt{h_B(i) h_W(i)}. \quad (2)$$

This distance has been selected as it is reported to have good properties [9]. We can make the following observations concerning its behavior in this context:

1. d_B can be treated as a measure of separation of the two histograms. In other words, it can be used to estimate the compactness of the clusters formed by realizations of each gesture. Good separation (preferable situation, with high d_B value) occurs when, on average, the distance between gestures sampled from one class is substantially smaller than to other classes. Poor separation occurs when the average distance within each class is comparable to other classes. In the second case the d_B value is low, and there can be expected a loss of performance due to misclassification of gestures.
2. d_B changes with m, as the number of symbols together with reference data define the quantization function, which produces symbol strings.
3. With a large value of m (large number of symbols), a situation can arise where a number of symbols will appear only once in single realizations (sequences). Hence the distances between sequences within-class will be large and comparable to those between-class, and d_B value will be low. This phenomenon is akin to overtraining of VQ.
4. With a small value of m, the pool of symbols will be small, hence all sequences will be similar (in terms of distance function). This can be viewed as "undertraining", and the result can be expected to be similar to above – distances comparable (although on average lower), and low d_B value.

Based on the observations, it seems fitting to observe the $d_B(m)$ and find its maximum, corresponding to best separation between the classes. While not formally proved, a number of experiments with different equipment and gesture sets confirm the validity of the above observations.

2.2 Choosing the Initial Parameters of HMM

For training the HMM, the Baum-Welch (BW) method is commonly used. It is noted [1,3,4], that the local optimization approach used in BW requires a good estimate of initial conditions, in particular of the symbol emission matrix B. One approach is to choose the number of states n with trial and error, training HMM for each choice with a number of randomized B candidates and choosing the combination providing maximum performance. This approach is essentially a random search in parameter space, and can be time consuming with a large number of models and/or states of each model, as the number of free parameters to be estimated is on the order of $O(n^2)$.

We propose to obtain initial estimates for HMM matrices for a given class using a hierarchical clustering of data sequences with the following steps:

1. Split the k training sequences into the l fragments each. Index the fragments as (t,i), $t=1,\ldots,l, i=1\ldots,k$ and compute the distribution of symbols in each fragment.
2. Generate candidate initial HMM (see below).

3. Perform reestimation and assess current performance (for example, empirical likelihood or Akaike information criterion [10])

4. Merge step: compute Bhattacharyya distances of distributions of symbols in fragments (a) for given t (b) between set of fragments at t and $t+1$. Merge the fragments (or sets) with smallest distance (most similar).

5. Repeat the process (goto 2) until either the desired performance target is satisfied, or until number of fragments is 1.

The initial HMM generation in step (2) is done as follows. Number of states is equal to the number of fragments. Matrix B is set with symbol distributions for each fragment. Matrix A and vector Π are initialized with left-to-right structure (with A matrix structured so that from each state i only the transfer to $j \geq i$ is possible). However, if, for given t, there is more than one fragment, it is assumed that different exclusive symbol distributions can appear (for example users will make the gesture with one of several hand configurations). This exclusivity is reflected in the state transition matrix structure: a state created from fragment (t,i) has non-zero probability of transfer only to itself and to all states $(t+1,\cdot)$, while having zero probability of transfer to all others, including $(t,j), j \neq i$.

3 Experiments and Results

We have performed experiments to observe and confirm validity of the proposed method. The data set used consists of motion capture recordings of 22 natural gestures ("A-OK", "thumbs up", etc) with four participants and five repetitions of each gesture per person (with different speeds: three times normal speed, one fast, one slow execution). The equipment used was DG5VHand motion capture glove, containing five resistance type finger bend sensors and three axis accelerometer producing three

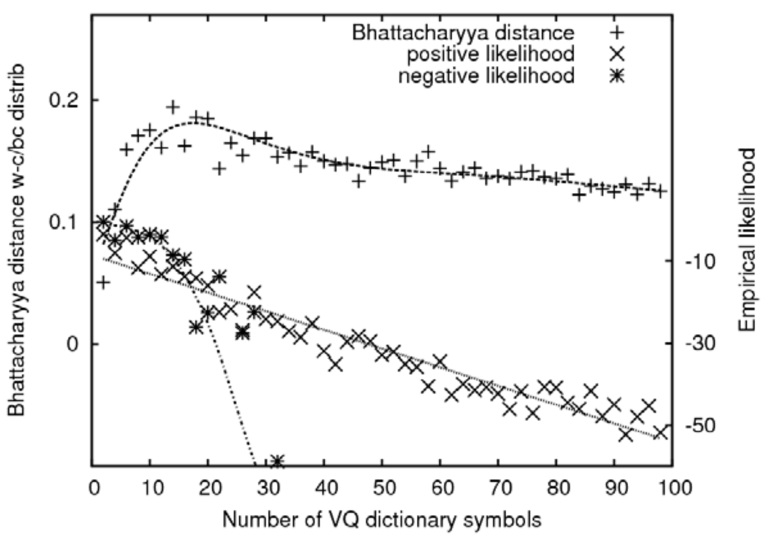


Fig. 1. Values of Bhattacharyya distance for different numbers of symbols, and log likelihoods: positive (of the class that model was trained on) and negative (of other classes)

acceleration and two orientation readings. The sampling frequency was ~30Hz. The movements used were a mix of natural, real-life gestures of several types (symbolic, iconic, deictic and manipulative), selected and captured with an objective on creating a diverse, rich testing set for natural gesture recognition methods. The length of recorded sequences was between 30 and 190 vectors. The experiment software was written in C++ using several publicly available numerical algebra libraries. For details on the creation of database and gesture choice, refer to [7].

For VQ dictionary the K-means algorithm was used with Mahalanobis distance (the covariance matrix was estimated from data). The range of symbols considered was m=2,...,200. The observed histograms of Levenshtein distance show noticeable overlap, mainly due to the fact that, with different execution speeds, the sequences for the same gesture differ in length (by 30-40% between fast/normal or normal/slow speeds). While Bhattacharyya distance estimates are noisy, the tendency to peak at m~18 can be seen (fig. 1). The behavior of HMMs for different number of symbols was then observed. First, one HMM was trained for each gesture (with number of states $n=20$; models chosen from l=10 candidates; left-to-right topology with uniform initialization for matrix A and vector Π; B matrix randomized with simplex method). Then, the log likelihood was computed – for given model i, of the true class i (positive) and $j \neq i$ (negative). As noted before, with small number of symbols, all sequences are similar, even across different gestures (small inter-class variation, hence high negative likelihood); with large number of symbols, all gesture realizations are dissimilar, even for the same gesture (large intra-class variation, decrease of positive likelihood). Proposed method selects maximum of Bhattacharyya distance, which corresponds roughly to the area of decline of negative likelihood.

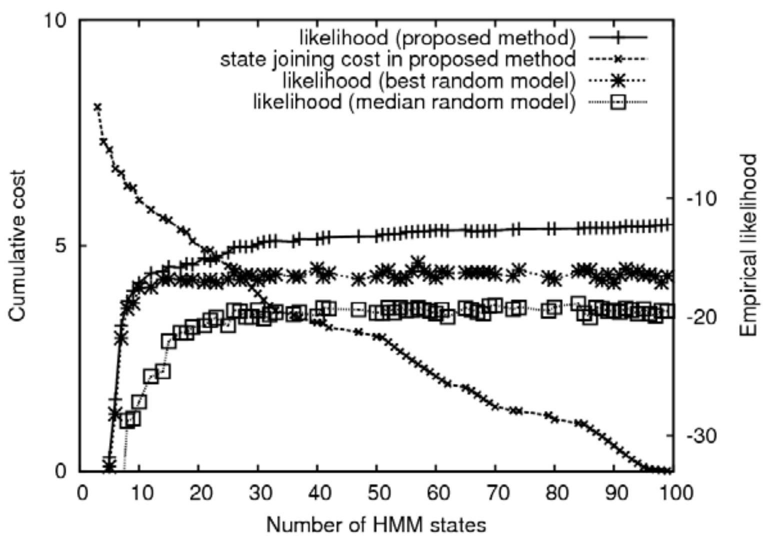


Fig. 2. Comparison of log likelihoods with proposed method and random initialization, with the cumulative cost of merging (sum of Bhattacharyya distances for subsequent fragment joining)

The second set of experiments was carried out to assess the performance of the fragment clustering method. For each number of states, one HMM was initialized with proposed method, and *l=100* randomly (with the same method as above). The log likelihoods of proposed method, minimum (best value) and median of random models are presented on fig 2. The improvement is increasing with number of states. The likelihood increase when *n>30* is not substantial, this can be used as a cue for *n* choice; for exact value, one can rely on selection methods such as [6].

4 Conclusions

We have presented an approach for unsupervised parameter selection of VQ+HMM classifier and experiments validating the algorithm on a real dataset of natural gestures. The main advantage of its application is the possibility to investigate the parameter space and find the best performing values with cost of several scales of magnitude smaller than exhaustive search. As the VQ+HMM is very common in the role of time-varying sequence classification, this approach has potentially wide range of applications, for example motion capture or camera based HCI systems employing gesture recognition.

Acknowledments. This work has been partially supported by a Polish Ministry of Science and Higher Education project NN516405137 "User interface based on natural gestures for exploration of virtual 3D spaces".

References

1. Rabiner, L.R.: A tutorial on Hidden Markov Models and selected applications in speech recognition. Proceedings of the IEEE 77(2), 257–286 (1989)
2. Lee, H.-K., Kim, J.H.: An HMM-Based Threshold Model Approach for Gesture Recognition. IEEE Trans. Pattern Analysis and Machine Intelligence 21(10), 961–973 (2002)
3. Brand, M., Oliver, N., Pentland, A.: Coupled hidden Markov models for complex action recognition. In: Proc. Computer Vision and Pattern Recognition (CVPR), pp. 994–999 (1997)
4. Nathan, K., Senior, A., Subrahmonia, J.: Initialization of hidden Markov models for unconstrained on-line handwriting recognition. In: IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), vol. 6, pp. 3502–3505 (1996)
5. Bicego, M., Murino, V., Figueiredo, M.A.T.: A sequential pruning strategy for the selection of the number of states in hidden Markov models. Pattern Recognition Letters 24(9-10), 1395–1407 (2003)
6. Celeux, G., Durand, J.-B.: Selecting hidden Markov model state number with cross-validated likelihood. Computational Statistics 23(4), 541–564 (2008)
7. Głomb, P., Romaszewski, M., Opozda, S., Sochan, A.: Choosing and modeling hand gesture database for natural user interface. In: Proc. of GW 2011: The 9th International Gesture Workshop Gesture in Embodied Communication and Human-Computer Interaction (2011)
8. Ouivirach, K., Dailey, M.N.: Clustering human behaviors with dynamic time warping and hidden Markov models for a video surveillance system. In: International Conference onElectrical Engineering/Electronics Computer Telecommunications and Information Technology (ECTI-CON), pp. 884–888 (2010)

9. Kailath, T.: The Divergence and Bhattacharyya Distance Measures in Signal Selection. IEEE Trans. on Communication Technology 15(1), 52–60 (1967)
10. Akaike, H.: A new look at the statistical model identification. IEEE Trans. on Automatic Control 19(6), 716–723 (1974)
11. Linde, Y., Buzo, A., Gray, R.: An Algorithm for Vector Quantizer Design. IEEE Transactions on Communications 28(1), 84–95 (1980)
12. Somervuo, P., Kohonen, T.: Self-Organizing Maps and Learning Vector Quantization for Feature Sequences. Neural Processing Letters 10(2), 151–159 (2004)
13. Martin, M., Maycock, J., Schmidt, F.P., Kramer, O.: Recognition of Manual Actions Using Vector Quantization and Dynamic Time Warping. In: Graña Romay, M., Corchado, E., Garcia Sebastian, M.T. (eds.) HAIS 2010. LNCS, vol. 6076, pp. 221–228. Springer, Heidelberg (2010)
14. Rigazio, L., Tsakam, B., Junqua, J.-C.: An optimal Bhattacharyya centroid algorithm for Gaussian clustering with applications in automatic speech recognition. In: Proc. ICASSP 2000 IEEE International Conf. on Acoustics, Speech, and Signal Processing, vol. 3, pp. 1599–1602 (2000)

Number Entry Interfaces and Their Effects on Error Detection

Patrick Oladimeji[1], Harold Thimbleby[1], and Anna Cox[2]

[1] Future Interaction Technology Lab, Swansea University, SA2 8PP, UK
[2] UCL Interaction Center, University College London, WC1E 6BT, UK
{p.oladimeji,h.thimbleby}@swansea.ac.uk, anna.cox@ucl.ac.uk

Abstract. A significant amount of interaction involves number entry. The purpose of any number entry interface is to accurately select or set a numeric value. There are two main styles of number entry interfaces found on medical devices: serial interfaces like the ubiquitous 12-key numeric keypad, and incremental interfaces that use a knob or a pair of keys to increase or decrease numbers. We report an experiment that investigates the effect of interface design on error detection in number entry. The initial findings show that the incremental interface produces more accurate inputs than the serial interface, and the magnitude of errors suggests that the incremental interface could reduce the death rate relative to the numeric keypad.

Keywords: number entry, data entry, error detection.

1 Introduction

Number entry is perceived to be a very simple and mundane task — yet numerical drug dosing errors account for a significant portion of adverse drug events (ADEs) in hospitals, particularly in paediatrics [2, 5]. Number entry errors can be as a result of a combination of user errors, poor interaction design and hardware defects. Hardware defects on keypads (e.g., key bounces) have been reported by the Institute for Safe Medication Practices (ISMP) as a source of error in medical device programming [1]. A key bounce occurs when physically pressing a key once causes a repeat of the same key; this is different from a double keying error where a user accidentally presses the same key twice. If a key bounce is undetected, it might lead to an overdose, which in turn might lead to patient harm or death.

In many user interfaces, number entry is implicit; for example, adjusting sound levels by rotating an unmarked dial, or moving a scroll bar adjusts a hidden number — but the user copes because of direct feedback (direct manipulation). In this paper we are concerned with numbers that are displayed by the user interface as precise Arabic numerals (e.g., 123, 6.5, etc) rather than as abstract values (e.g., colors), or even as numbers indicated by pointers on numeric scales (e.g., analog meters). This style is prevalent in safety-critical applications such as drug rate control in healthcare.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 178–185, 2011.

Number entry interfaces may be grouped into two main categories:

- **Serial number entry:** A familiar example of serial entry is the 12-key numeric keypad used on calculators. The user enters the desired number typically using a keypad. The number is entered serially, digit by digit starting from the leftmost digit. There are two main variations: those with a decimal point key and those without. When there is no decimal point key, a decimal point is inserted in a fixed position, typically giving 2 decimal digits (so keying 1234 would input 12.34). Serial entry provides the quickest method for arbitrary number entry, and its use also coincides with how numbers are spoken in many Western languages.
- **Incremental number entry:** In an incremental interface, the user is only able to increment or decrement the number, using fixed actions, typically a pair of up/down keys, rotary dials, or pressure sensitive controls. As a result, the number entry is about changes to a current number. The user may be able to influence the rate of change of the number by exerting more pressure on the control (in the case of a pressure sensitive interface) or by the speed with which the control is manipulated (e.g., turning a knob), or by the length of time the interaction is prolonged for (e.g., holding down an increment key longer to invoke faster or larger changes on a number). User interfaces typically refine these styles, for instance to impose numerical limits and by varying key layout.

Unfortunately, errors are eventually inevitable when using interactive systems and these can be in the form of mistakes, slips or lapses [7]. Sometimes, errors are detected by users and corrected. When errors go undetected, the consequences can be very serious. In a safety critical and dependable system, it is important that users realise when they commit errors and correct the errors. We believe that the differences in the design of number entry user interfaces place different demands on users in terms of what part of the user interface they focus most of their attention on and as a result whether they notice that an error has occurred and correct the error.

In this paper, we investigate the behaviour of users performing a routine task of number entry using serial and incremental user interfaces. We report our findings on the relative accuracies of both styles and a classification of the types of errors committed by users.

2 Experiment

Because interaction on the incremental interface requires users to monitor how the value on the display changes based on what key the user is interacting with, we hypothesize that users will more likely detect and correct errors when using an incremental interface as opposed to when using a serial interface. We also anticipate that users will pay more visual attention to the display than to the input when using the incremental interface and pay more visual attention to the input than to the display when using the serial interface. Finally, we expect that there will be types of errors that are unique to each class of interface.

Fig. 1. Screen shot of the incremental interface tested in the experiment

2.1 Participants

22 participants (12 female) aged 18–55 years took part in the experiment. All participants were regular users of computers. None were prone to repetitive strain injury. One was dyslexic. Each participant was compensated for their time with a gift voucher.

2.2 Apparatus

A computer with an integrated Tobii eye-tracker was used to present the instructions and the number entry interfaces. Interaction with the computer was with a mouse, which was used to click on "keys" on the screen. 100 numbers were generated randomly for the experiment with the following constraints: all numbers were between 0 and 10, all numbers had a decimal point, all numbers had at least one non-zero significant digit after the decimal point and all numbers were different.

The serial interface was based on the Graseby 500 infusion pump. It allowed number entry using a full numeric keypad in the calculator style layout. This interface had a decimal point key and had a cancel key for deleting the rightmost character on the display. This interface allowed a maximum of 5 characters in its display, which may include only one decimal point.

The incremental interface was based on the Alaris GP infusion pump. It had four keys (Fig. 1.). The first two keys, with the upward pointing chevrons, increased the value displayed and the last two keys decreased the value. For each of the two sets of keys, one key (with the double chevron) caused a bigger change while the other caused a change that is ten times smaller. This interface allowed two modes of interaction. The user could click the keys or press and hold the keys. Clicking the keys changed the displayed value as specified above. Pressing and holding the keys changed the displayed value at a rate proportional to the duration the key was held down for. Users were expected to press and hold for faster increments or decrements. This interface always showed a decimal point and two digits after the decimal point.

A key bounce was triggered for the 84th, 88th, 92nd and 97th entry for both interfaces to see if users will detect and correct the errors. The experiment logged mouse actions to obtain accuracy and performance data on the number entry tasks.

2.3 Design

The experiment was a within-subject repeated measures design. Each participant used both number entry interfaces. The number entry interface style was the independent variable and it had two levels: the incremental and serial interfaces. The order in which the interfaces were tested was counterbalanced for all participants. The dependent variables were the number of undetected errors, number of corrected errors, total eye fixation time on the input and display part of the interface and task completion times.

2.4 Procedure

All participants were tested individually. The eye tracker was calibrated for each participant and they were briefed about the stages and purpose of the experiment before starting.

The experiment itself was in two parts: one for each interface. Prior to each part of the experiment, the participant got a training session where they could enter 10 numbers and get familiar with the interface. When the participant was comfortable with how the interface worked, they were allowed to proceed to the experiment.

For the experiment, each participant was required to enter 100 numbers using both interfaces in the order defined by the experimenter. The participants were instructed to enter the numbers as quickly and as accurately as possible. An instruction on the right half screen showed what number the participant should enter. The participant had to click a 'Next' key to confirm their entry, and display the next instruction. The process of number entry and confirmation of entry was repeated until all 100 numbers had been entered using the first interface. The participant was allowed a break of up to 5 minutes before proceeding to the second half of the experiment. The interface was then switched and the participant went through the training session for that interface and proceeded to enter the same set of 100 numbers.

2.5 Results

Four participants were excluded from the statistical analyses. One participant's error rate was more than two standard deviations from the mean of error rates, and three had very low eye tracking data. Undetected errors were instances of number entry errors that were not caught and corrected by the participants before confirming the entry. These were not limited to the four key bounce errors introduced in the experiment. All but two participants had at least one undetected error in the experiment. To check for learning effects on the interfaces over the experiments, we analysed the task completion times from each interface in blocks of 10 trials per participant and we found no significant difference in the mean time per block.

Corrected errors for each participant on the serial interface were calculated as the total number of times they pressed the 'Cancel' button. For the incremental interface, the corrected errors for each participant were calculated as the number of times the participant overshot or undershot the target number. In the incremental interface, overshooting the target number was sometimes intentional especially when entering numbers efficiently using a mixture of continuous and discrete actions. For instance, entering the value '5.9' efficiently means slightly overshooting the target number using the continuous (hold-down) interaction in order to reach '6', for instance, and refining the value with a down click (a discrete action) to obtain the target number. This type of intentional overshooting did not count as a corrected error.

Undetected Errors

Analysis of the total undetected errors for both interfaces using a paired t-test indicated that the mean undetected errors for the incremental interface (mean=1.56, sd=1.76) was significantly lower than that of the serial interface (mean=3.61, sd=2.38), $t(17)=-4.15$, $p<0.001$.

The total undetected forced errors per participant on the incremental interface (mean=0.28, sd=0.46) was significantly lower than that of the serial interface (mean=1.72, sd=1.23), $t(17)=4.74$, $p<0.001$.

Corrected Errors

The number of corrected errors per participant on the incremental interface (mean=74, sd=17.31) was significantly greater than the number of corrected errors in the serial interface (mean=7.1, sd=9.6), $t(17)=17.22$, $p<0.001$.

Visual Attention

The total visual fixation duration on the input of the serial interface (mean=271.16secs, sd=80.01), was significantly greater than the total fixation duration on the display of the device (mean=26.28secs, sd=19.31), $t(17)=13.35$, $p<0.001$. Conversely, the total fixation duration on the input of the incremental interface (mean=185.82secs, sd=87.78) was significantly lower than the fixation duration on the display (mean=553.47secs, sd=276.25), $t(17)=7.34$, $p<0.001$.

Speed of entry

The speed of entry per trial for the incremental interface (mean=8.2secs, sd=2.32) was significantly slower than the speed of entry per trial for the serial interface (mean=1.65secs, sd=0.33), $t(17)=12.71$, $p< 0.001$.

Error Types

Keystroke logs from all participants were analysed for this section. Below, we report a selection of the undetected error types that occurred in our experiment. Wiseman, Cairns and Cox have developed a taxonomy of number entry errors [6] and have independently reported and classified these errors. As well as reporting error types, we report the prevalence of certain error types between the two number entry interface styles. The frequency of each error type is shown in Figure 2. For each error type we quantify the severity of the errors by reporting the mean and standard deviation of the difference between the intended number and the transcribed number.

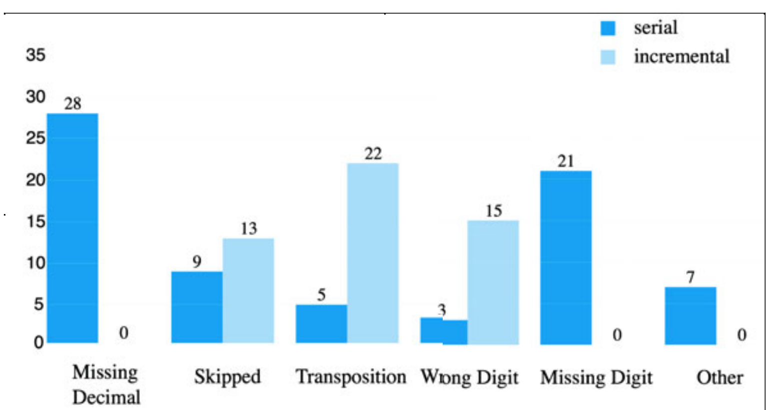


Fig. 2. The distribution of undetected error types in each interface in the experiment. Skipped errors were due to double keying errors on the "Next" key during the experiment

Missing Decimal Point Errors
This error occurs when a decimal point is absent from the transcribed number but is present in the instruction. There were 28 instances of this error on the serial interface and none on the incremental interface. On average, this error changed the intended number by 260.77 (sd=240.85).

Transposition Errors
Transposition errors occur when the user swaps two adjacent digits in a number. For instance instead of entering 5.84, a user might enter 5.48. The majority of these were committed on the incremental interface. The dyslexic participant committed no transposition errors. Most of the transposition errors occurred after the decimal point.

In our data, a special case of the transposition error occurred when the decimal part of the transcribed number was exactly 10 times more or less than the decimal part of the intended number. For example, instead of entering 7.4, a participant entered 7.04. Although one participant committed 17 of these errors, the potential causes make it a concern for further investigation. It is possible that the display of the numbers on the incremental interface was responsible for this error: the display always shows three significant digits. For instance if the numeric value is 7.4, the display shows 7.40. It may be confusing that the 40 after the decimal point is perceived to be greater than 4. It is important to note that this participant did not commit any transposition errors on the serial interface. It seemed that the incremental interface had an effect on their transcription of numbers specifically for numbers of the form 'd.0d' where 'd' is a numeric digit. In other words, the interface design might have affected their perception of numbers of a certain format.

Transposition errors were more serious on the incremental interface. On average this error changed the intended number by 0.54 (sd=0.35) on the incremental interface compared to 0.31 (sd=0.18) on the serial interface.

Wrong Digit Errors
Wrong digit errors occur when one of the digits in the transcribed value is incorrect. This error was more common in the incremental interface. The most serious cases of the wrong digit error happened whenever the whole number part of the number is wrong. For instance a participant entered 4.87 instead of 5.87. Wrong digit errors were more serious on the serial interface but more frequent on the incremental interface. On average, this error changed the intended number by 0.81 (sd=1.27) on the serial interface compared to 0.28 (sd=0.40) on the incremental interface.

Missing Digit Error
This refers to instances of errors where one digit from the intended value is missing from the transcribed value. For instance a participant entered 0.3 instead of 0.43. On average, this error changed the intended number by 3.36 (sd=8.92). The incremental interface was free of this error.

3 Discussion

The results show a significantly higher number of undetected errors on the serial interface in comparison to the incremental interface. This relative accuracy however comes with a slower data entry speed on the incremental interface, which may in itself account for the more accurate performance on the incremental interface. The higher level of visual attention paid to the display of the incremental interface is another possible reason for its higher accuracy. A third reason could be that participants expect to make errors on this interface. Indeed, the results show a significantly higher number of corrected errors on the incremental interface in comparison to the serial interface. Some participants had a number of tries overshooting and undershooting for the intended number before precisely setting the number. Some deliberately overshoot the intended value and correct the error in a few clicks because that is the optimal way to enter the intended number.

For the incremental interface, the visual attention placed on the input was significantly lower than that on the display. This supports our original intuition, as the interaction on the incremental interface requires the user to monitor how the value on the display changes based on what key the user is pressing. The input part of the incremental interface requires little visual attention and is only used to switch direction and precision of change. However, despite the high attention paid to the display of this interface, the mode of interaction introduced errors that were less likely on the serial interface e.g., the wrong digit errors and the transposition errors.

The results also show that the visual attention placed on the input in the serial interface was significantly higher than the visual attention on the display. This could be because participants did not feel the need to verify their entry. It is possible that most participants trusted the visual feedback they got from the labels on the keys and felt little need for an extra mode of feedback by checking the display.

By design, the numbers specified on a serial interface require parsing to obtain a numeric value valid in the application space. As a result, serial interfaces are prone to syntax errors. Rather than alert users to errors, this parsing process often produces incorrect and unpredictable results whenever the user commits a syntax error [3, 4].

Syntax errors are however impossible on an incremental interface since the application guides the user through a valid range of numeric values. It is also plausible that numbers are perceived as a string of characters when using a serial interface whereas using an incremental interface forces users to be aware of the numeric values and the relative order of numbers.

In a safety critical context like healthcare, the incremental interface is safer. It allows better error detection and the severity of errors is much lower than on the serial interface. The missing decimal point and the missing digit errors are the most serious errors and they were both more likely to occur on the serial interface. Overall, the errors on the incremental interface had a much lower deviation from the intended number as shown by the difference of intended value to transcribed value (serial interface: mean=70.91, sd=166.94, incremental interface: mean=0.93, sd=1.41).

4 Conclusions

There are significant differences in the error rates for the two experimental conditions of number entry: number entry interface styles do affect error rates — and, by implication, medical outcomes. The speed of the serial interface comes at a price: errors are more likely to go undetected due to significantly less visual attention on the interface and undetected errors like the missing decimal or missing digit are more likely to have serious outcomes typically producing numbers out by a large factor (10 or more) from the intended values. In a medical context, such errors can be fatal.

The result suggests that it should be a priority to research number entry styles and their relation to error rates, behaviour and performance. There is a wide variety of number entry styles in medical devices (where errors cause adverse events), clearly with no or little empirical justification; we now see useful progress can be made to provide sound guidance for designers of safety critical number entry systems.

Acknowledgements. Funded as part of CHI+MED: Multidisciplinary Computer-Human Interaction research for design and safe use of interactive medical devices project EPSRC Grant Number EP/G059063/1. Duncan Brumby provided invaluable feedback on this paper.

References

1. ISMP: Double key bounce and double keying errors. In: ISMP Medication Safety Alert! Acute Care (January 2006), http://www.ismp.org/newsletters/acutecare/articles/20060112.asp (accessed April 2011)
2. Jani, Y.H., Barber, N., Wong, I.C.K.: Paediatric dosing errors before and after electronic prescribing. Quality and Safety in Health Care 19(4), 337–340 (2010)
3. Thimbleby, H.: Interactive systems need safety locks. In: ITI 2010: The 32nd International Conference on Information Technology Interfaces, pp. 29–36 (2010)
4. Thimbleby, H., Cairns, P.: Reducing number entry errors: solving a widespread, serious problem. Journal of the Royal Society Interface 7(51), 1429–1439 (2010)
5. Vicente, K.J., Kada-Bekhaled, K., Hillel, G., Cassano, A., Orser, B.A.: Programming errors contribute to death from patient-controlled analgesia: case report and estimate of probability. Canadian Journal of Anesthesia 50(4), 328–332 (2003)
6. Wiseman, S., Cairns, P., Cox, A.: A Taxonomy of Number Entry Error. In: HCI 2011: The 25th BCS Conference on Human-Computer Interaction (2011)
7. Reason, J.: Human Error. Cambridge University Press, Cambridge (1990)

An Exploration of the Utilization of Electroencephalography and Neural Nets to Control Robots

Dan Szafir and Robert Signorile

Computer Science Department,
Boston College, Chestnut Hill, MA USA
{szafird,signoril}@bc.edu

Abstract. It has long been known that as neurons fire within the brain they produce measurable electrical activity. Electroencephalography (EEG) is the measurement and recording of these electrical signals using sensors arrayed across the scalp. The idea of Brain-Computer interfaces (BCIs), which allow the control of devices using brain signals, naturally present themselves to many extremely useful applications including prosthetic devices, restoring or aiding in communication and hearing, military applications, video gaming and virtual reality, and robotic control, and have the possibility of significantly improving the quality of life of many disabled individuals. The purpose of this research is to examine an off the shelf EEG system, the Emotiv EPOC© System, as a cost-effective gateway to non-invasive portable EEG measurements and to build a BCI to control a robot, the Parallax Scribbler®. We built middleware to interpret the outputs from the Emotiv and map them into commands for the Scribbler robot.

Keywords: Human-Robot Interaction, Computer Human Interface, Control Systems, Neural networks.

1 Introduction

Simple Brain-Computer interfaces (BCIs) currently exist and research and public interest in them only continues to grow. This research explores the process in creating a novel BCI that utilizes the Emotiv EPOC System to measure EEG waves to control the Parallax Scribbler robot. We wrote middleware to interpret the signals from the Emotive system and map those signals into commands for the Scribbler robot. These include commands such as move forward, move to an obstacle, etc. In latter sections of this paper, we describe the current research in BCI, describe the Emotiv system, the Scribbler robot, our middleware and a novel use of blinking (which can be detected by the Emotive systems) to enhance the command set for the robot [29].

2 Electroencephalography

EEG waves are created by the firing of neurons in the brain and were intensely in the fields of neuroscience and psychology [1] [2]. EEG waves are measured using

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 186–194, 2011.

electrodes attached to the scalp, which are sensitive to changes in postsynaptic potentials of neurons in the cerebral cortex. Postsynaptic potentials are created by the combination of inhibitory and excitatory potentials located in the dendrites. These potentials are created in areas of local depolarization or polarizations following the change in membrane conductance as neurotransmitters are released. These average of the potentials are amplified and combined to show rhythmic activity that is classified by frequency [3]. Electrodes are usually placed along the scalp as in Figure 1 [4]. One of the historical downsides of EEG measurement has been the corruption of EEG data by artifacts, which are electrical signals that are picked up by the sensors that do not originate from cortical neurons. One of the most common causes of artifacts is eye movement and blinking, however other causes exist [5]. Many EEG systems attempt to reduce artifacts and general noise by utilizing reference electrodes placed in locations where there is little cortical activity and attempting to filter out correlated patterns [6].

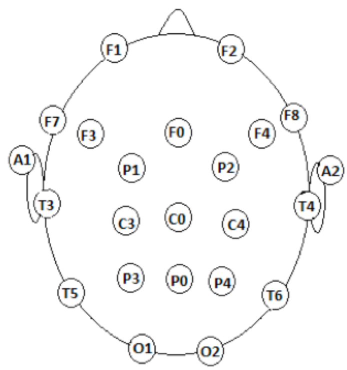


Fig. 1. Electrode Placement according to the International 10-20 System. Odd numbers on the right, even on the left. Letters correspond to lobes – F(rontal), T(emporal), P(arietal), and O(ccipital). C stands for Central (there is no central lobe).

3 Brain-Computer Interfaces

The term "Brain-Computer Interface" is the idea of linking the mind to computers [7]. The ultimate goal of BCI research is to create a system that is not only an "open loop" system that responds to users thoughts but a "closed loop" system that also gives feedback to the user. Researchers initially focused on the motor-cortex of the brain, the area which controls muscle movements, and testing on animals quickly showed that the natural learning behaviors of the brain could easily adapt to new stimuli as well as control the firing of specific areas of the brain [8] translate them into robotic activity [9][10] [11] [12]. Research is beginning to veer away from invasive BCIs due to the costly and dangerous nature of the surgeries required for such systems. Non-invasive alternatives for BCIs include EEG technology, Magnetoencephalography (MEG), and Magnetic Resonance Imaging (MRI), as well as the "partially invasive" Electrocorticography where sensors are placed within the skull but outside the gray matter of the brain. These methods are limited in that they are often susceptible to noise, have worse signal resolution due to distance from the brain, and have difficulty recording the inner workings of the brain. However, non-invasive techniques have the advantage of lower cost, greater portability, and the fact that they do not require any special surgery [13].

4 Previous EEG BCI Research

Though the idea of using EEG waves as input to BCIs has existed since the initial conception of BCIs, actual working BCIs based on EEG input have only recently appeared [14]. Most EEG-BCI systems follow a similar paradigm of reading in and analyzing EEG data, translating that data into device output, and giving some sort of feedback to the user (Figure 2), however implementing this model can be extremely challenging [15]. The primary difficulty in creating an EEG-based BCI is the feature extraction and classification of EEG data that must be done in real-time if it is to have any use.

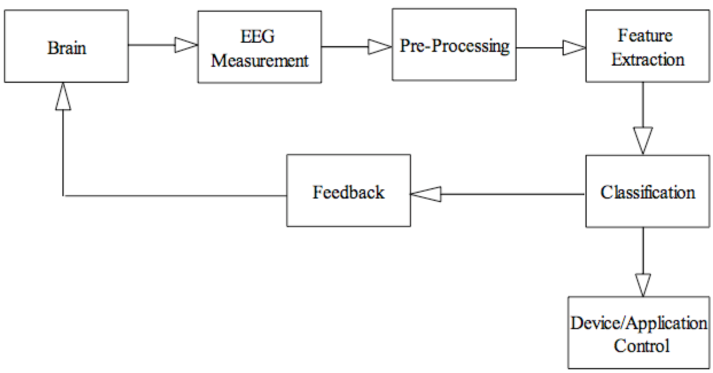


Fig. 2. Brain-Computer Interface Design Pattern

Feature extraction deals with separating useful EEG data from noise and simplifying that data so that classification, the problem of trying to decide what the extracted data represents, can occur. There is no best way of extracting features from EEG data and modern BCIs often use several types of feature extraction including Hjorth, wavelet transforms and Fourier transforms. The major features that EEG-BCI systems rely on are event-related potentials (ERPs) and event-related changes in specific frequency bands [16][17]. BCI systems are further complicated by the fact that there is no standard way of classifying the extracted data. Various types of pattern recognizers are employed to try to match the input data to known categories of EEG archetypes [18]. Researchers have also relied on unsupervised learning algorithms to find natural clusters of EEG segments that are indicative of certain kinds of mental activities with varying degrees of success [19][20]. Feedback is essential in BCI systems as it allows users to understand what brainwaves they just produced and to learn behavior that can be effectively classified and controlled [21].

EEG-BCIs can be classified as either synchronous or asynchronous. The computer drives synchronous systems by giving the user a cue to perform a certain mental action and then recording the user's EEG patterns in a fixed time-window. Asynchronous systems are driven by the user and operate by passively and continuously monitoring the user's EEG data and attempting to classify it on the fly. Synchronous protocols are far easier to construct and have historically been the primary way of operating BCI systems [22] [23][24].

5 Our Research Project: Combining Machine learning, Neural Nets, Brain Waves and Our Middleware to Control Personal Robots

The goal of our project was to investigate and suggest the use of brainwaves to control personal robots (and thus demonstrate the more general proposition that robots can be controlled by brainwaves). Using training of the Emotiv System, we were able to extract the EEG signals from the headset, categorize them into one of several groups, translate that group to a robotic command, and finally control the robot. The two major hardware components of our system are the Emotiv Headset and the Scribbler robot. The next subsections briefly describe each system. Then we discuss our integration of the two systems sand finally we describe some adjustments we made to extend the overall system.

5.1 The Emotiv© System

The Emotiv System is based around the EPOC headset for recording EEG measurements and software suit which processes and analyzes the data. This research originally uses the Research Edition of this off the shelf product. The Research Edition includes the Emotiv Control Panel, EmoComposer (an emulator for simulating EEG signals), EmoKey (a tool for mapping various events detected by the headset into keystrokes), TestBench, which enables the capture of raw EEG data from each individual sensor [26].The Emotiv system can measure engagement/boredom, frustration, meditation, instantaneous excitement, and long-term. The Cognitiv suite can measure 13 active thoughts as well as the passive neutral state. This software works by running the input from the electrodes through a neural network and attempting to classify the signals as one of the 13 built-in “prototype” action thoughts.The core of the Emotiv SDK is the “EmoEngine,” which is a logical abstraction that “communicates with the Emotiv headset, receives preprocessed EEG and gyroscope data, manages user-specific or application-specific settings, performs post-processing, and translates the Emotiv detection results into an easy-to-use structure called an EmoState.” Every EmoState represents the current input from the headset including “facial, emotional, and cognitive state” contains electrode measurements for each contact. Utilizing the Emotiv API consists of connecting to the EmoEngine, detecting and decoding new EmoStates, and calling code relevant to the new EmoState (Figure 3).

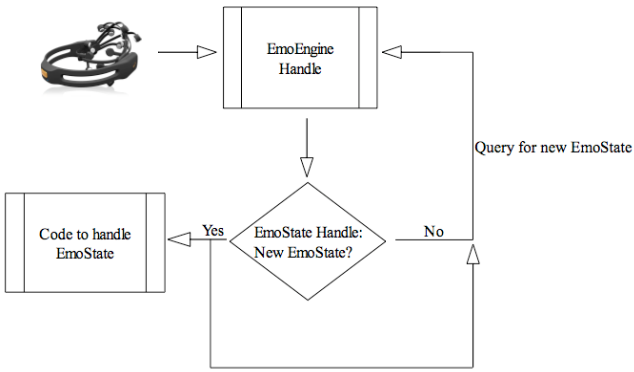


Fig. 3. High-level View of the Utilization of Emotiv

5.2 The Parallax Scibbler® Robot and IPRE Fluke

The Parallax Scribbler robot is a fully assembled reprogrammable robot built around the BASIC Stamp® 2 microcontroller. It contains built in photovoltaic sensors, infrared sensors, line sensors, two independent DC motors to drive the two wheels, three LED lights, a speaker, and a serial port[28]. The Institute for Personal Robots in Education (IPRE) Fluke is an add-on board created by Georgia Robotics that plugs into the Scribbler's serial port and adds color vision, IR range sensing, internal voltage sensing, an extra LED, and bluetooth functionality and has created the Myro APIs to control the Scribbler using Python [29].

5.3 Control Implementation

The code implementing this control scheme is divided into four basic parts: connecting to the Emotiv headset via the Emotiv API, connecting to the Scribbler through the

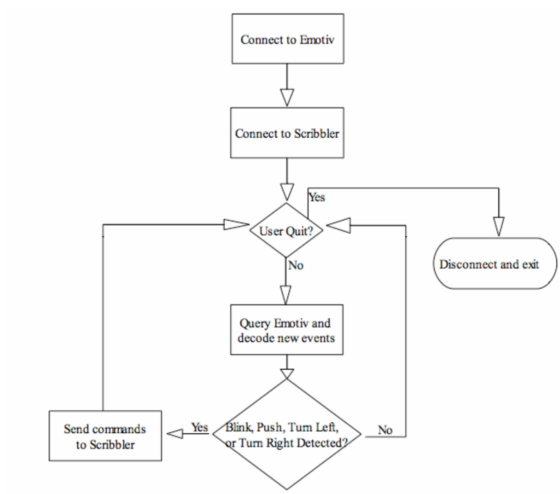


Fig. 4. High-level View of the Control Scheme

Myro Python libraries, reading and decoding Emotiv events and sending the corresponding commands to the Scribbler, and closing the connections when the user is done (Figure 4).

5.4 Decoding and Handling EmoStates

There are four major steps in reading and decoding information from the EPOC headset: creating the EmoEngine and EmoState handles, querying for the most recent EmoState, deciding if this is a new EmoState, and decoding the EmoState. The EmoEngine handle allows for queries to get direct input from the headset including contact quality, raw electrode input, and the connection quality. New EmoStates are constantly created by the EmoEngine which represent recognized actions such as facial expressions, changed emotional status, and detected thoughts and can be queried through the EmoState handle.Next the program determines the event type returned by the EmoEngine. There are three categories of event types: hardware-related events, new EmoState events, and suite-related events. Once we have decoded which thought sparked the EmoState, we send the appropriate call to the Scribbler (Push → Move Forward, Turn Left → Turn Left, Turn Right → Turn Right). We initially experimented with using the power of the thought as an input to the power of the robotic action, however we found that this control scheme was too difficult to use and it ended up being far more intuitive to use specific values for turning and moving forward no matter the thought-power. This allowed the user to concentrate on the thoughts alone and not have to additionally worry about how "hard" to think about the thoughts.

The internal sampling rate in the EPOC headset is 2048Hz which is filtered to remove artifacts and alias frequencies own to about 128Hz. Any given motion input to the Scribbler using Bluetooth takes approximately 2 seconds to execute, while picture taking take slightly longer as it has to capture and send data back. Thus we needed to synchronize the sampling rate with the actual Scribbler command execution (so that commands are not just queued up, but executes in a orderly fashion). To solve this problem, we introduced a sampling variable to only decode one in every ten input EmoStates to limit the EmoStates created. Using this sampling variable we filter out those extra states that really only correspond to one event by using a sample rate small enough that it will still capture events which send more than 10 input EmoStates while sending only one command to the Scribbler instead of queuing up 10. This system worked much better, and even had the added bonus of filtering out "noise" thoughts when the headset detected a push or turn thought for a fraction of a second.

To extend the number of command we can send to the Scribbler, we created an additional mode, which remaps the same input thoughts to different outputs in the robot. This is hugely beneficial as it does not increase the difficulty in recognizing new thoughts and also does not require the user to train additional thoughts, thus giving double the usability with only one additional input. This additional input is raising the eyebrows, which toggle between the original and the new mode. We decided on utilizing the raising of eyebrows as a toggle as it is very easily trained and accurately recognized by the headset. The addition of more modes is certainly possible and is an excellent way of adding functionality without adding the cost of recognizing and learning new thoughts. In the end, it was completely feasible to control the Scribbler robot using the EPOC headset proving the viability of EEG based BCI technology.

5.5 Blink Detection and Data Reduction/Noise Reduction

We next decided to explore the Pre-Processing, Feature Extraction, and Classification of EEG data by analyzing eye blinks. The analysis of eye blinks is useful in BCI development for two reasons: eye blinks can be used as control inputs and if they are not they must be filtered out lest they corrupt the useful EEG data. We decided on eye blinks since they are immediately recognizable in one particular channel from the headset. This allowed us to immediately reduce the amount of input data by a factor of 14 since we could discount the other 13 input channels.

Reducing the size of the data set is the primary focus of pre-processing and feature extraction whose goal is to get rid of extraneous and noisy data while preserving data that can best differentiate between classes. We recorded twenty, ten second clips, ten of which we blinked during and ten of which we didn't. We then exported the clips to MATLAB. These recordings produced a lot of data because in addition to the 14 EEG channels capturing electrode data at 128Hz the headset also records gyroscope data, battery data, packet information, etc. and each 10 second clip ended up had roughly 36000 data points and combined I recorded 720000 data points.

The first step of our feature extraction was to use just the channel where blinks are clearly visible. The classification using MATLABS' nprtool to create a two-layer feedforward neural network with backpropagation obtained only a 65% accuracy. Though there was a certain pattern to the blinks, the neural net was thrown off because the blinks were not normalized with respect to time. The neural net treated time as an attribute, and thus did not classify two samples that both contain blinks but where the blinks occur at different times. Time was correlated to blinks in respect to how long the blink takes and thus how wide the blink spike will be, however the time that the blink occurs is not a usable attribute.

To solve this problem, we decided to further reduce the amount of data the neural net worked with along with normalizing any blinks found. Recognizing that blinks correlate to spikes in EEG data, we scanned each 10-second clip looking for the largest spikes. We found that blinks typically were represented by a surge in the 4500 to 4800 μvolt range over approximately .59 seconds and were followed by a characteristic dip of around 50 μvolts over approximately .20 seconds. This pattern was very clear and easily distinguishable from a non-blink state; we first noticed it when applying unsupervised K-Means Clustering to detect naturally occurring patterns in the data.We used this information to further filter each of the 20 inputs down to 1.6-second segments, each of which highlighted the maximum spike of the original ten-second segment. This normalized the blinks by having each blink start at roughly the same time and additionally filtered out noise that was unrelated to the blinks creating data that was much easier for a neural net to distinguish. Using these inputs, neural net accuracy improved to 100%, however we wanted to see if this system truly recognized blinks or was over-trained on the input data. We then recorded five more segments, 3 with blinks and 2 without, and followed the same pre-processing/feature extraction steps and fed the data to the neural net. The neural net accurately predicted all of these new inputs even though it had not been trained upon them, showcasing that it was truly was extendable and actually recognizing blink patterns. These are very promising results that prove the feasibility of utilizing a neural net to classify blinks, however it would be best to obtain a larger sample size to accurately test the classification performance of this scheme. Now blinks could be used to effectively double our thoughts (think of the recognizable blink as a “shift” key on a keyboard).

6 Conclusions

Part of our research was to examine the state of EEG-based BCI construction and implementation. In particular, we wanted to test the feasibility of BCI to control personal items such as a personal robot. Our investigation demonstrates that, with our middleware, it is a feasible system that will likely only improve and become more widespread in the future. The system we constructed was largely a success as we were able to create a system whereby we could control a robot with our thoughts and we further created accurate blink-recognizing software to enhance the amount of thoughts we can recognize.Furthermore, our system showcases the possibilities of BCI's in aiding the disabled. For instance, a person who could only move their head could certainly use our system to control a motorized wheelchair accurately using their thoughts. In addition, had they a computer built into the headset, they could easily switch modes by raising their eyebrows and then use their thoughts as an input to the computer, by using the same thoughts that had moved the wheelchair to control the mouse and double-blinking to click. An alternative would be to keep the thoughts controlling the wheelchair while utilizing the gyroscope in the headset to control the mouse, enabling the user to have simultaneous control of the wheelchair and computer. Further research can certainly lead to direct implementation of such systems and can explore the recognition of thoughts beyond those included in the Emotiv API.

References

1. Swartz, B.E., Goldensohn, E.: Timeline of the history of EEG and associated fields. Electroencephalography and Clinical Neurophysiology 106, 173–176 (1998)
2. Millett, D.: Hans Berger: from psychic energy to the EEG. Perspectives in Biology and Medicine 44(4), 522–542 (2001)
3. Nunez, P.L., Srinivasan, R.: Electric Fields of the Brain: The Neurophysics of EEG. Oxford University Press, Oxford (1981)
4. Niedermeyer, E., da Silva, F.L.: Electroencephalography: Basic Principles, Clinical Applications and Related Fields, 5th edn., p. 140. Lippincott Williams & Wilkins (2005)
5. Rowan, A.J.: Primer of EEG. Elsevier Science, Philadelphia (2003)
6. Ludwig, K.A., et al.: Employing a Common Average Reference to Improve Cortical Neuron Recordings from Microelectrode Arrays. Journal of Neurophysiology, September 3 (2008)
7. Vidal, J.: Toward Direct Brain–Computer Communication. Annual Review of Biophysics and Bioengineering 2, 157–180 (1973)
8. Fetz, E.E.: Operant Conditioning of Cortical Unit Activity. Science 163, 955–958 (1969)
9. Kennedy, P.R., et al.: Activity of single action potentials in monkey motor cortex during long-term task learning. Brain Research 760(1-2), 251–254 (1997)
10. Wessber, J., et al.: Real-time prediction of hand trajectory by ensembles of cortical neurons in primates. Nature 408(6810), 361–365 (2000)

11. Carey, B.: Monkeys Think, Moving Artifiacl Arm as Own. The New York Times, May 29 (2008)
12. Hochberg, L.R., et al.: Neuronal ensemble control of prosthetic devices by a human with tetraplegia. Nature 442, 164–171 (2006)
13. Fabiani, G.E., et al.: Conversion of EEG activity into cursor movement by a brain-computer interface
14. Vidal, J.: Toward Direct Brain–Computer Communication
15. Omidvarnia, A.H., et al.: Kalman Filter Parameters as a New EEG Feature Vector for BCI Applications
16. Niedermeyer: Electroencephalography, pp. 1265–1266
17. Sellers, E.W., et al.: A P300 event-related potential brain–computer interface (BCI): The effects of matrix size and inter stimulus interval on performance. Biological Psychology 73(3), 242–252 (2006)
18. Adlakha, A.: Single Trial EEG Classification. Swiss Federal Institute of Technology, July 12 (2002)
19. Lu, S., et al.: Unsupervised Brain Computer Interface based on Inter-Subject Information. In: 30th Annual International IEEE EMBS Conference, Vancouver, British Columbia, Canada, August 20-24 (2008)
20. Niedermyer: Electroencephalography, p. 1240
21. Kauhanen, L., Palomaki, T., Jylanki, P., Aloise, F., Nuttin, M., Millan, J.R.: Haptic feedback compared with visual feedback for BCI. In: Kauhanen, L., Palomaki, T., Jylanki, P., Aloise, F., Nuttin, M., Millan, J. (eds.) Proceeding of 3rd International BCI Workshop and Training Course 2006, Graz, Austria, September 21-25, pp. 66–67 (2006)
22. Niedermeyer: Electroencephalography, p. 1265
23. Birbaumer, N., et al.: The Thought Translation Device (TTD) for Completely Paralyzed Patients. IEEE Transactions on Rehabilitation Engineering 8(2), 190–193 (2000)
24. Galán, F., et al.: A Brain-Actuated Wheelchair: Asynchronous and Non-invasive Brain-Computer Interfaces for Continuous Control of Robots. Clinical Neurophysiology 119(9), 2159–2169 (2008)
25. Drummond, K.: Pentagon Preps Soldier Telepathy Push. Wired Magazine, May 14 (2009)
26. Emotiv Website, `http://www.emotiv.com`
27. The Scribbler: A Reprogrammable Robot, `http://www.parallax.com/tabid/455/Default.aspx`, Copyright 2010 by Parallax Inc. (accessed November 4, 2010)
28. Usage Guides, `http://www.roboteducation.org/guides.html`, Copyright © 2007 Institute for Personal Robots in Education (accessed November 4, 2010)
29. Szafir, D.: Non-Invasive BCI through EEG, unpublished Undergraduate Honors Thesis in Computer Science, Boston College (2010)

Social Translucence as a Theoretical Framework for Sustainable HCI

Mary Barreto, Evangelos Karapanos, and Nuno Nunes

Madeira Interactive Technologies Institute,
University of Madeira, Caminho da Penteada,
Funchal, 9020-105, Madeira, Portugal
{maryluizbarreto,e.karapanos}@gmail.com, njn@uma.pt

Abstract. Motivating sustainable behaviors is increasingly becoming an important topic in the HCI community. While a substantial body of work has focused on the role of peer-pressure through social networks, we argue that the community has largely overlooked the importance of strong social ties and specifically those of family members. We propose the theory of Social Translucence as a theoretical framework for understanding how eco-feedback interfaces can integrate with and support existing communication practices within families. We report on our ethnographic inquiry involving a day reconstruction study followed by in-depth interviewing with 12 families, which took place during a six-month deployment of an eco-feedback interface. Through our study we attempt to inquire into how eco-feedback interfaces: a) raise mutual awareness of family members' consumption behaviors, and b) induce feelings of accountability on individuals regarding their consumption behaviors.

1 Introduction

Environmental sustainability is an increasingly important global issue. Energy consumption, in particular, is considered to be invisible to millions of users, which makes a prime case of much wastage especially in domestic environments [3]. New technologies in data acquisition and analysis (for instance Non-Intrusive Load Monitoring) provide individuals access to information about their energy consumption that is otherwise difficult to estimate. This has lead to an outburst in interest in the so-called *eco-feedback technology*, "technology that provides feedback on individual or group behaviors with a goal of reducing environmental impact" [6].

An increasing body of work on eco-feedback technologies stresses the influences that social networks exert on individuals' behaviors. Social influences may contribute towards sustainable behaviors through stimulating competition and providing social incentives [15], supporting public goal commitment [14] and affecting social norms of a culture [7]. In one of the first examples of such work within HCI, Mankoff et al. [12] discussed opportunities and concerns when leveraging the power of social networking sites in influencing individuals' actions. Odom et al. [13] developed and evaluated different eco-visualizations for use in student dormitories. Stepgreen [12] is a social networking site that enables individuals to assess the impact as well as

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 195–203, 2011.

receive feedback on their goals with respect to sustainable behaviors. Last, even commercial services such as Google powermeter and Microsoft ohm have adopted social features.

Most existing work has focused on weak social ties, such as neighbors, friends or contacts in online social networking sites. These types of social ties are, however, expected to exert weaker influence on individuals' consumption behaviors when compared with stronger social ties such as family relations [4, 8]. Families often discuss about and encourage pro-environmental behaviours, they maintain a higher awareness of each others' behaviours, and display limited privacy concerns relating to in-house activities and whereabouts, when compared to weaker social ties [4, 8].

This paper attempts to contribute to our understanding of how families appropriate eco-feedback interfaces in their daily routines. We employ the framework of Social Translucence as a theoretical lens and we raise two questions: a) how eco-feedback interfaces raise mutual awareness of family members' behaviors, and b) induce feelings of accountability on individuals regarding their consumption behaviors.

2 Eco-feedback Interfaces as Socially Translucent Systems

The Theory of Social Translucence [5] argues that motivating desired behavior requires more than making one's behavior visible to his or her social network. It identifies three properties – *visibility*, *awareness*, and *accountability* – of socially translucent systems, systems that support coherent behavior in groups and communities by making participants and their activities visible to one another.

Socially translucent systems first have to make socially significant information, such as one's energy consumption or transport behavior, visible to one's social network. Once this information is visible, people may or may not become aware of this and may act upon it. For instance, they may positively respond to a good act and thus reinforce it or may also become motivated to behave in the same manner. Thirdly, this mutual awareness of each other behaviors eventually results into people feeling more accountable for their actions.

Visibility refers to making one's behavior (e.g. energy consumption) visible to others. We understand visibility in a broad sense, reflecting eco-feedback interfaces' ability to make not only family behaviors, but also the impact of those behaviors, visible among all members of the family. In other words, eco-feedback interfaces need to visualize consumption behaviors within a house but also challenge family misbelieves about what actions may result to energy savings.

According to Social Translucence theory, however, visibility does not guarantee *awareness* of the information. Contextual aspects such as the location of the eco-feedback interface, as well as the role of different family members when it comes to energy consumption may influence the extent to which families maintain awareness of each others' behaviors. Is the eco-feedback interface accessible to all members of the family, or does the family use one or some of its members as a proxy to the information? Second, mutual awareness of each other's consumption behaviors exists even in the absence of eco-feedback interfaces; the question, then, is: how do eco-feedback interfaces leverage existing communication practices of the family rather than replace them?

Last, the theory of Social Translucence postulates that *accountability* of one's consumption behavior is built up through the mutual awareness of each other's actions. Through making consumption behaviors visible to all family members, eco-feedback interfaces are expected to impact the social structure of families. The question raised is: do eco-feedback interfaces participate harmonically in families, and in what ways do they induce feelings of accountability on family members' behaviors?

3 Study

3.1 Participants

Participants were selected from a list of energy consumption data given by the local electricity company. We selected two buildings that displayed high variation in the energy consumption across individual apartments and invited families to participate in the study through leaving letters in the mailboxes. Out of the 35 apartment occupants that responded to the invitation we selected 30 families who showed interest in the study, and which belonged to the four annual energy consumption levels defined by the company. Two out of the 30 families refused to participate in the study when the system was about to be installed, as they argued it would not be safe to have it around their children. The current study presents findings from our interviews with 12 of these 28 families. Families' residence in those apartments ranged from 2 to 11 years (μ=6.25, sd=2.59). Parents' age ranged from 30 to 48 years for the fathers (μ=37.83, SD=5.75), 29 to 46 years old for the mothers (μ=35.25, SD=5.78). 10 out of the 12 families had children. Three children were of age 0-2, four of age 4-5, six of age 7-11, three of age 13-15, and one of age 26.

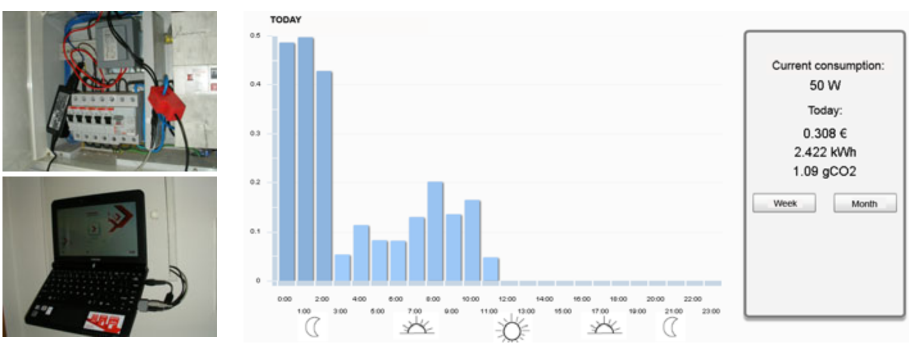


Fig. 1. (left) the powermeter consisted of a netbook and an ADC converter, next to the mains fuse box which in all participants' apartments was located in the main corridor (right). The eco-feedback interface presented information relating to the household's overall consumption per day, week or month, and in terms of KWh, cost and CO_2 emissions.

3.2 Eco-feedback as Technology Probe

To sensitize families on energy consumption we introduced a technology probe [9], an energy meter with a simple eco-feedback interface. Our interest was to understand how this system would probe discussion within the family and how families would appropriate it within their daily routines. The energy meter consisted of a netbook and an ADC converter for measuring household energy consumption and inputting it to the netbook through the microphone input. The netbook was installed next to the mains fuse box, which in all participants' apartments was located in the main corridor, a place that we judged appropriate for a public display (see figure 1). The eco-feedback interface presented information relating to the household's overall consumption per day, week or month, and in terms of KWh, cost, and CO_2 emissions (see figure 1). The interface was built in Adobe Flex while all computations and handling of user events (i.e. mouse activity, recording people either passing by or standing in front of the netbook) were performed in Processing.

3.3 Method

The study consisted of two parts: a) a 1-day diary of all members of the family followed by b) interviews the day after. Using the Day Reconstruction Method (DRM) [10], we asked all family members to list, in a chronological order, the activities they performed while being within the house during the reported day. For each activity they provided a brief name and start and end time. Following the complete reconstruction of all daily activities, participants were asked to provide more detailed information for each activity. This information was: a) electrical devices that were directly or indirectly used in the activity, b) all family members' locations (kitchen, living room, bedroom, outside house, other part of house), and c) subjective ratings on the three social translucence dimensions, namely visibility, awareness and accountability using 7-point likert scales.

Interviews, the day after, took place with all family members present as we wanted to understand how family members would interact when addressing energy consumption issues. Interviews consisted of three parts. Firstly, in a warm-up discussion we probed for general information relating to the family's daily routine, their concerns and their expenses with utilities. Secondly, each member of the family was asked to go through the diary and select one or two activities that they would like to discuss. For each activity we asked them to reminisce the context and motivations for the performed activities and probed for discussion between her and other members of the family on whether and how this activity could be improved with respect to energy consumption. Thirdly, we conducted a contextual inquiry in the usage of the behavior meter and its implication on their behaviors and concluded with a summary of the most substantial insights while asking how the reported day (in the diary study) is different from a typical family day. Qualitative data were analyzed using Affinity Diagrams [1]. Individual statements were printed and posted to a wall. They were then clustered in hierarchical themes followed by labeling the themes. This was an iterative process performed by the first two authors.

4 Findings

Similarly to Broms et al. [2] and Karapanos et al. [11], we observed two phases in families' interactions with the eco-feedback interface: an initial *orientation* phase that lasted approximately 4 weeks, and after which we saw a decline of about 40% in users' interactions, followed by an *incorporation* phase which signified, first, the loss of the powermeter's novelty, and secondly its appropriation in families' daily routines. In the remainder of the paper we describe our findings with regard to our two research questions:

4.1 How Do Eco-feedback Interfaces Raise Mutual Awareness of Family Members' Behaviors?

Our probe revealed to make consumption behaviors more visible to the family either through making these more transparent, bringing them to debate or by challenging their performance when conducting them.

Through supporting peripheral awareness. During the incorporation phase the powermeter was primarily used for *reassurance* purposes, i.e. knowing that everything is as one would expect them to find. The mains fuse box where the powermeter was placed was often located in the main corridor of the apartment and next to its main entrance. Glancing at the powermeter provided peripheral awareness of each others' behaviors both in terms of *attention resources* and *time*. Individuals could infer the devices being used at a single point in time only based on the overall – disaggregate – household consumption. This often cued an inference about others family members' behaviors such us in the following example: *"I saw high consumption and went around to see that the fridge's door was open … they always forget to close it properly!"* (family 7 Father, ref 1). In other cases, the powermeter allowed absent family members to infer household activities on their arrival: *"One day I used the dryer to dry some clothes… [my husband] arrived home and asked what I had used between those hours as the consumption was 3 times higher. I said nothing much and then I remembered I had used the dryer"* (family 4 Mother, ref1).

Through cueing discussion. We were surprised to observe that only a limited set of cues (such as disaggregate household consumption per hour) was enough to allow rich inferences about household activities. This was possible partly because families do maintain awareness of each other's activities as one father shared with us, *"we must coordinate all our tasks, so it is very rare that either she or I have any consumption that the other doesn't know what it is"* (family 19, Father, ref7). But even more importantly, we found that, when background information was not sufficient to allow inferences about who and what consumed energy, the powermeter provided cues for discussion among family members as in the previous example of family 4. We noticed that depending on the type of the activity, families displayed different co-presence patterns (operationalized in terms of number of members *collocated* in the same room as well as number of members outside the house), with high-consuming activities like doing the laundry or ironing often being solitary (see table 1). Families' activities seem to also have a different distribution across time during weekdays and during weekends (see Figure 2). For instance, activities such as personal care, cooking and having a meal happen around the same time during weekdays but are more dispersed during weekends. TV watching was a constant activity throughout both weekdays and weekends. Activities integrated in leisure time were more prone to happen during the weekend.

Through supporting arguments with data. Families revealed to hold misbelieves about each others' consumption behaviors and often conflicts arose, e.g. *"he constantly switches on the TV whenever he walks around the house, even if he's not there watching, in the kitchen, in the bedroom, in the living room..."* (family 7, Mother, ref3). The powermeter gave family members the ability to support their arguments with data, as in one case where children complained about the fathers' use of his personal computer in response to his criticism on their use of console games. In other cases, family members used energy cues to infer behavior, such as when parents observed that their children spent too much time with console games, or the presence of the maid: *"on Mondays the maid comes and vacuums, cleans, irons as we checked in the computer [powermeter] that day is an energy peak for us"* (family 28, Father, ref1). In fact, family members rated activities higher in terms of energy consumption when these meant using high perceived consuming devices (cooking, and laundry) or when they would spend a considerable amount of time performing these activities (console games) (see table1).

Fig. 2. Distribution of family activities over time during weekdays and weekends

4.2 How Do Eco-feedback Interfaces Induce Feelings of Accountability on Individuals Regarding Their Consumption Behaviors?

As it became obvious in the previous section, the powermeter did not provide any radically new information to the family, but instead, it came to enhance the presence of consumption information: *"After seeing how much using the dryer during the day costs, I am more aware and I will avoid using it. One thing is getting the bill at the end of the month and I see numbers, the other is seeing it there on the screen" (family 4 Mother, ref1).*

Through leveraging families' existing means. Families have their own means for inducing accountability in individuals' consumption behaviors, such as commenting on others' behaviors, adapting one's own behavior to set the example, leaving subtle messages (e.g. a parent placing environmental magazines at a visible location), or even employing creative ways to do so, e.g. *"I use some tape in the switch so they don't use it every time they come*" (family 26, Mother, ref 1).

The powermeter leveraged those means through enhancing the presence of costs and environmental impact of energy consumption, both in making it present throughout the whole month (in contrast to the monthly bill) but also in making it accessible to all members of the family, e.g. *"They only know [how much it costs] when we complain about the bill, when it is too high"* (family 4 Mother, ref 1). For instance, parents would use energy consumption data in educating their children on pro-environmental behavior and expense management: *"I had to explain to her (to the youngest daughter) why we need to pay it (electricity) (…) sometimes she wants to stay more time in the water and I tell her we need to save water and gas* (family 26 Mother, ref 4). In other cases the powermeter supported more frequent and data-grounded reflection on the family's energy expenses: *"I talk to my husband… it is just to keep track of it, we mostly talk about it we don't write it down"* (family 5 Mother, ref 1).

Table 1. Number of family members collocated (in the same room), number of family members being outside the house, perceptions of energy consumption and perceived accountability for 13 activity categories. Mean values (standard deviations).

Activities	*No Collocated*	*No Outside*	*Energy*	*Accountability*
Have the meal	2,64 (1,32)	0,42 (0,77)	4,36 (3,10)	4,95 (2,42)
Cooking	1,50 (0,74)	0,53 (0,88)	5,61 (4,22)	4,43 (2,85)
Personal care	1,83 (1,17)	0,19 (0,49)	2,57 (2,89)	4,66 (2,51)
Care for others	2,50 (0,85)	0,60 (0,84)	2,70 (3,27)	2,60 (3,21)
School time	2,25 (1,49)	1,00 (0,92)	1,75 (0,89)	4,50 (0,71)
Work time	1,73 (1,32)	1,20 (1,15)	4,33 (2,61)	4,87 (2,77)
Cleaning	1,57 (0,79)	0,14 (0,38)	3,86 (2,03)	3,57 (3,41)
Washing dishes	2,50 (0,71)	1,50 (2,12)	3,50 (2,12)	4,50 (0,71)
Doing laundry	1,16 (0,41)	0,16 (0,41)	5,00 (3,69)	6,00 (1,09)
Ironing	1,00 (0)	0,33 (0,58)	4,33 (1,53)	6,66 (0,58)
Watching TV	2,32 (1,25)	0,61 (1,08)	3,81 (1,66)	4,64 (2,18)
Console games	1,33 (0,52)	1,50 (1,64)	4,50 (0,55)	1,00 (0)
Leisure time	1,77 (0,83)	0,85 (0,89)	3,15 (1,95)	4,00 (1,53)

Through enforcing transparency - I know that you know. Surprisingly, participants' ratings on accountability over the different reported activities in the day reconstruction study displayed a low correlation with ratings of perceived energy consumption (Pearson's $r = 0.12$, $n = 240$, $p = 0.05$). Contrary, perceptions of accountability correlated with perceptions of awareness suggesting that one feels more accountable when other family members know about his or her energy consumption (Pearson's r = 0.59, n = 240, $p < 0.01$). Activities such as having a meal or watching TV (see table 1), which were largely shared activities, displayed higher ratings of accountability. We found that the powermeter changed the way families communicated about consumption behaviors. Firstly, it motivated all members to share their knowledge when inferring what activities caused particular consumption levels as it supported a common family goal. Secondly, through raising mutual awareness of each other behaviors, it induced perceptions of accountability both during engaging with energy consuming behaviors as well as during reasoning and discussing over individuals' behaviors with family members.

5 Conclusion

This paper proposed the theory Social Translucence as a framework for understanding how eco-feedback interfaces raise mutual awareness of, and, secondly, induce feelings of accountability on individuals' consumption behaviors. We reported on our initial interviews with 12 families during our 6-month deployment of a simple eco-feedback interface. Our future work will attempt to further inquire into family dynamics using different forms of probing and will attempt to draw implications for the design of eco-feedback interfaces.

References

1. Beyer, H., Holtzblatt, K.: Contextual design: defining customer-centered systems. Morgan Kaufmann Pub., San Francisco (1998)
2. Broms, L., Katzeff, C., Bang, M., Nyblom, A., Hjelm, S.I., Ehrnberger, K.: Coffee maker patterns and the design of energy feedback artefacts. In: Proceedings of the 8th ACM Conference on Designing Interactive Systems, pp. 93–102 (2010)
3. Darby, S.: The effectiveness of feedback on energy consumption. Environmental Change Institute, University of Oxford (2006)
4. Edgerton, E., McKechnie, J., Dunleavy, K.: Behavioral determinants of household participation in a home composting scheme. Env. and Beh. 41(2) (2009)
5. Erickson, T., Kellogg, W.A.: Social Translucence: An Approach to Designing Systems that Support Social Processes. ACM TOCHI 7(1), 59–83 (2000)
6. Froehlich, J.: Promoting energy efficient behaviors in the home through feedback: The role of human-computer interaction. In: HCIC 2009 Winter Workshop Boaster Paper (2009)
7. Goldstein, N., Cialdini, R., Griskevicius, R.: A Room with a Viewpoint: Using Social Norms to Motivate Environmental Conservation in Hotels. J. Consumer Research 35 (2008)
8. Haeffner, M., Casalegno, F.: How Does a Visual Monitoring System Foster Sustainable Behavior? Instructional Technology, 27–35 (2009)
9. Hutchinson, H., Mackay, W., Westerlund, B., Bederson, B.B., Druin, A., Plaisant, C., Beaudouin-Lafon, M., Conversy, S., Evans, H., Hansen, H., et al.: Technology probes: inspiring design for and with families. In: CHI 2003, pp. 17–24 (2003)
10. Kahneman, D., Krueger, A.B., Schkade, D.A., Schwarz, N., Stone, A.A.: A survey method for characterizing daily life experience: The day reconstruction method. Science 306(5702), 1776–1880 (2004)

11. Karapanos, E., Zimmerman, J., Forlizzi, J., Martens, J.-B.: User Experience Over Time: An Initial Framework. In: Proceedings of the 27th International Conference on Human Factors in Computing Systems, CHI 2009, pp. 729–738. ACM, New York (2009)
12. Mankoff, J., Fussell, S.R., Dillahunt, T., Glaves, R., Grevet, C., Johnson, M., Matthews, D., Matthews, H.S., McGuire, R., Thompson, R., et al.: StepGreen. org: Increasing energy saving behaviors via social networks. In: Proceedings Fourth International AAAI Conference on Weblogs and Social Media, pp. 23–25 (2009)
13. Odom, W., Pierce, J., Roedl, D.: Social Incentive & Eco-Visualization Displays: Toward Persuading Greater Change in Dormitory Communities. In: Workshop Proc. of OZCHI, p. 8 (2008)
14. Pallak, M.S., Cummings, W.: Commitment and voluntary energy conservation. Personality and Social Psychology Bulletin 2(1), 27–30 (1976)
15. Petersen, J.E., Shunturov, V., Janda, K., Platt, G., Weinberger, K.: Dormitory residents reduce electricity consumption when exposed to real-time visual feedback and incentives. International Journal of Sustainability in Higher Education 8(1), 16–33 (2007)

A Revised Mobile KLM for Interaction with Multiple NFC-Tags

Paul Holleis[1], Maximilian Scherr[2], and Gregor Broll[1]

[1] DOCOMO Euro-Labs, Landsberger Str. 312, 80687 Munich, Germany
{holleis,broll}@docomolab-euro.com
[2] Ludwig-Maximilians-Universität München, Amalienstr. 17, 80333 Munich, Germany
scherr@cip.ifi.lmu.de

Abstract. The Keystroke-Level Model (KLM) is a model for predicting the execution time of routine tasks. Initially, it had been devised for standard keyboard-desktop settings but an extension of this model for interactions with mobile phones has been described by Holleis et al. [10]. We propose a considerable update of this KLM focusing on NFC-based applications and interactions which are continuously gaining interest. Insufficiencies within the previous model regarding operators for Pointing, Mental Acts, and System Response Time are treated. We present the results of several studies conducted in order to update the values of these operators. A specific focus is put on the differences between static (NFC tags behind a printed poster or object) and dynamic interfaces (tagged displays or projections). Finally, we validate our results by modeling two applications with the former and the proposed model. The latter performed consistently better when compared with measurements from real user interaction data.

Keywords: Keystroke-level model (KLM), mobile device interaction, Near Field Communication (NFC), static / dynamic multi-tag interaction.

1 Introduction

The Keystroke-Level Model (KLM) [5] is an established tool for assessing the performance of application designs in early phases of development before prototyping or implementation. It can predict the average execution time for specific tasks by modeling how much time expert users need to accomplish them on the level of basic input events, e.g. moving the mouse or pressing a key. For that purpose, KLM defines a set of *operators*, e.g. pointing, keystrokes or mental preparation, as well as average execution times for them that have been derived from experiments. The added execution times for KLM operators of a task can then be used to assess or compare different application designs for the same task.

Originally, KLM has been developed to assess the performance of desktop applications that get input from a keyboard and a mouse. With the advancement of Mobile Computing, it has also been applied and adapted to mobile devices and applications. Mobile devices enable new kinds of interactions that use Bluetooth, cameras, GPS, Near Field Communication (NFC) [25] or other technologies to

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 204–221, 2011.

interact with the real world and its objects. In order to incorporate these and other aspects of mobile interaction, Holleis et al. [10] have introduced the Mobile KLM with updated operators and execution times of the original KLM, e.g. Keystrokes or Pointing. It also includes new operators to model attention shifts, distraction or interactions with objects that are tagged with NFC or visual markers [21].

In this paper, we revisit the Mobile KLM in order to bring it up to date with current technology and knowledge for mobile interactions with NFC-tags. NFC is a wireless technology for data exchange over short distances. Passive tags can be attached to almost arbitrary objects and can store digital data which can be read by touching them with NFC-enabled mobile devices. In the Mobile KLM, Holleis et al. [10] (see the detailed description below) only regard simple interactions with *single tags*. Although this is very convenient for mobile payment, ticketing or information retrieval, mobile interaction with NFC can be more elaborate. Everyday objects such as posters or advertising columns can comprise *multiple NFC-tags* and serve as physical user interfaces (UI) that adopt features of mobile applications, complement mobile UIs and support new interaction techniques. Additionally, we focus on novel types of applications that use *dynamic interfaces*. In contrast to *static interfaces* such as tagged objects or posters, dynamic applications use tagged displays or tagged screens on which an interface is projected.

In the following section, we summarize related work about (mobile) KLM and mobile interaction with NFC, focusing on multi-tag interactions. Based on this overview, we point out shortcomings of the Mobile KLM regarding system response time, pointing and mental preparation for mobile interaction with multiple NFC-tags. Next, we describe a comprehensive user study whose different parts investigate these shortcomings one by one. The results are combined into an update of the Mobile KLM for multi-tag interaction with NFC. Finally, we validate our findings in an additional study that compares estimated task execution times for multi-tag applications derived from the original Mobile KLM and its suggested update. The latter performed consistently better on all instances compared to measurements from real user interaction data.

2 Related Work

The Keystroke-Level Model [5] can be seen as a lightweight version of the GOMS (Goals, Operators, Methods and Selection rules) model that has been developed by Card et al. [6] to model error-free interaction with interfaces. KLM simplifies the original GOMS by providing a common set of operators and modeling only one method thus removing the need for selection rules. The resulting KLM can be used to predict the time it would take expert users to complete a task with the given interface. The basic KLM operators defined in [5] are *Pointing* (e.g. with a mouse or other pointing devices), *Keystrokes* (e.g. pressing keys or buttons), *Homing* (moving the hand between the mouse and the keyboard), *System Response Time* and *Mental Act* (i.e. mental preparation for a subtask). The KLM also provides average execution times for its operators that have been derived from several experiments. Those have been applied and adapted for different kinds of applications and interactions in order to provide reliable performance assessments for them.

Over the last 25 years, the KLM has proven to be effective in a great variety of application areas such as email organization [1], map digitization [9], navigation systems [16], or interfaces for people with disabilities [14]. Mobile devices, applications and interaction have also become an area of application for KLM taking into account the specific properties and constraints of this platform. Small screens and keyboards, context-awareness, and mobility make changes and additions in the existing KLM essential. One example can be found in Luo and Siewiorek [15] who extend it to also predict the energy consumption of interactions on mobile devices. Still, most of the early works on models for mobile devices and applications only assess simple interactions like navigating mobile phone menus [17]. Still, the model and its operators have already been extensively used and validated, see, e.g., Holleis [11] for an overview on existing KLMs for text entry systems on mobile devices.

In order to be able to describe and examine more advanced interactions on mobile phones, Holleis et al. [10] performed an extensive set of studies and provide an updated Mobile KLM. It revisits operators of the original KLM and adds new operators modeling novel types of interaction such as gestures or using tagged objects. The operators set in bold are of particular interest for the paper at hand. NFC interaction has been considered with the operators R (using 2.58s as response time for NFC) and A (used to model steps such as focusing the camera; it is set to zero for NFC interactions).

Table 1. Overview on KLM operators from Holleis et al. [10]. The ones set in bold or their uses are updated within this paper.

Operator	Function	Value	New?
A, Action	Certain interactions require additional actions	0.00s-1.23s	●
F, Finger	Finger movement on a small screen	0.23s	●
G, Gestures	Gesture interaction with finger, hand, etc.	0.80s	●
H, Homing	Moving phone to/from ears	0.95s	
I, Initial Act	Find and start interaction with device	1.18s-5.32s	●
K, Keystroke	Press a key / hotkey on the device	0.39s/0.16s	
M, Mental Act	Mental preparation for a subtask	1.35s	
P, Pointing	Moving the mobile device	1.00s	
R, System Response	Waiting for the system to respond	variable	
S_{Macro}	Attention shift world↔device	0.36s	●
S_{Micro}	Attention shift keypad↔display on device	0.14s	●
X, Distraction	Distraction by environment	6-21% slowdown	●

Since 2007, when the Mobile KLM was introduced, mobile interaction with tagged objects has considerably evolved. In particular, Near Field Communication (NFC) [25] has become a popular technology for mobile payment, ticketing, information retrieval or service interaction [20]. Apart from these simple interactions with single tags, NFC can also be used for more elaborate interactions with multi-tagged objects that have not been considered in the Mobile KLM. Applications can map features and options to multiple NFC-tags on physical objects that serve as physical UIs and thus complement mobile UIs. Examples are posters for mobile ticketing [2], tagged maps [19] or control panels for multimedia players [22].

In the next step, physical UIs are covered with a grid of NFC-tags to provide a completely interactive surface for the manipulation of dynamic application UIs on displays or projected interfaces. Opposite to static posters, these interactive surfaces are more flexible and can dynamically map their tags to the UI elements of different applications. Vetter et al. [24] and Hardy et al. [8] have created and refined the first dynamic NFC-displays that use a grid of NFC-tags for the interaction with an application UI that is projected onto this interactive surface. Similarly, Ramírez-González et al. [18] have combined a grid of NFC-tags and a projected application UI to build an interactive NFC-panel. Seewoonauth et al. [23] use a grid of tags on the back of a laptop display to enable direct, touch-based interactions.

3 Shortcomings of the Current Mobile KLM

The Mobile KLM of Holleis et al. [10] has proven valuable for the analysis of several novel mobile interactions. However, there are several shortcomings that hinder the exact modeling of NFC-based interaction, especially with multiple tags. From our experience with modeling and studying multi-tag interfaces, we identified several aspects that need to be improved:

1. System response time reading NFC tags with current mobile devices
2. Model errors caused by mismatches between tag-matrix and UI widgets
3. Pointing time for close and far objects
4. Chunking of action sequences needs to be re-evaluated

Item (1) is a common issue of models such as KLM. Although the general type and technique of interaction may remain the same across device models, hardware and system response time are bound to change. The time to complete a task can significantly change (in most cases decrease) when upgrading to a newer type of device or hardware. For NFC-based applications, the time for reading the content of a tag has a significant impact on the total task completion time and needs to be updated from the previous Mobile KLM.

Issues (2) to (4) arise from the fact that mobile interaction with NFC has changed compared to the applications envisioned in the previous Mobile KLM. The switch from single-tag to multi-tag interactions means that additional unit tasks appear and that the properties of some types of interaction changed. Item (2) has not been encountered before as there was always one UI element matched onto one tag. Some multi-tag applications use a grid of tags that covers the whole UI area. This allows for a more flexible use of the tagged area as UI widgets can be displayed everywhere and be rearranged dynamically. This also means that the direct mapping between tags and UI items is lost. Thus, a displayed button can span two or more tags and the user might touch the gap between adjacent tags, failing to read a tag and thus to press the button.

Item (3) dwells on the fact that the previous Mobile KLM suggests using an average of about one second to model any pointing operation. As proposed in the original KLM by Card et al. [5], this should be replaced by a dynamic model such as Fitts' Law [7] to make the model more precise. However, in practice this is often too cumbersome if it cannot be automated. Thus, constant approximations are often used.

In the case of multi-tag interaction, there are often many occurrences of a pointing action and smaller errors can add up more quickly than in single-tag interactions.

Item (4) refers to the placement of a special operator within the KLM: the Mental Act operator M is used to model some mental preparation that a user needs to perform before engaging into a physical action. However, certain sequences of actions can be anticipated and are not interrupted by a mental act. Guidelines for finding such sequences have been given by Kieras [13] and Holleis et al. [10]. These guidelines need to be updated for multi-tag interactions that have not yet been considered.

4 Studies to Update the Mobile KLM

In order to update the Mobile KLM for multi-tag interaction with NFC, we have conducted a comprehensive user study whose different parts focus on the previously elicited shortcomings.

1. Familiarization: At the beginning of the study, the subjects were introduced to mobile interaction with NFC. In order to become more familiar with this new technology, we let them play a game that uses a dynamic setup with a grid of NFC-tags and a projector that displays the game area on top of it.
2. Pointing: Moving the mobile device close to a tag is one of the most frequent actions for multi-tag interfaces. In the Mobile KLM, this is modeled with the Pointing operator P. This study mimics a Fitts' Law study by comparing pointing between targets of varying sizes and distances, both on static and dynamic UIs.
3. Sequences: The effect of consecutive touching of different tags is the subject of the second study. More precisely, the placement of Mental Act operators – describing brief pauses in the interaction when users need to remember or look for something etc. – is examined.
4. Measurement: The system response time for the most current NFC-enabled Nokia phone (6212) is measured and the respective value from the Mobile KLM is updated.

4.1 Study Setup and User Familiarization

An important aspect of working with models such as KLM is that only expert, routine behavior can be modeled with high accuracy. Although some extensions can also treat learning effects, see e.g. [12], investigators should only observe interactions where users are sufficiently familiar with the respective application. Errors should be kept to a minimum and the principle sequence of actions should be clear to the subjects without needing to check instructions, ask, or spend significant time thinking about a solutions strategy (unless, of course this should be part of the study or model).

In total, 16 subjects (13 male) with an average age of 30.3 years took part in the study. Three subjects did not have a background in computer science or related technical fields. On a Likert-scale from 1 ("none") to 5 ("expert"), the subjects rated their overall technical expertise and their expertise with mobile device as rather high (3.94 and 3.81). Although 9 of the 16 subjects had heard of NFC before the study, only 3 of them had actually used it.

In order to bring all subjects to roughly the same 'expert' level for the study, we explained its overall procedure and different parts. At the beginning of each part, we explained which aspect of mobile interaction with NFC it tested and let the subjects gain hands-on experience with the used NFC-enabled mobile phone and multi-tag UI. The subjects also performed each particular task of the study various times until they felt comfortable and understood the goal of the task and the method to reach it.

Before we started the actual studies, we had them play a prototype of the popular Whack-a-Mole game that we had previously adopted for dynamic NFC-displays [3]. This prototype projects the gaming area onto a grid of NFC-tags. Players have to touch its tags with an NFC-enabled mobile phone in order to hit the moles that appear on the gaming area and to win credits. The technical setup comprises a wall mounted grid of 48 by 20 Mifare NFC-tags, and a laptop running a Java SE game server that projects the game interface onto the tag grid via a ceiling-mounted short-throw projector. Players use a J2ME client application running on NFC-enabled Nokia 6212 phones. Each subject played two times alone for preparation and one multi-player round with one of the investigators. At this point, we conjecture that all subjects were familiar with the general concept of NFC tags, interaction techniques using a mobile phone and such tags, as well as the technology's shortcomings: reading errors due to gaps and overlapping tags in the tag mesh.

4.2 NFC System Response Time

The Mobile KLM [10] has already incorporated time estimates for detecting and reading an NFC-tag with an NFC-enabled mobile phone. However, this measurement has been done some years ago and only with the NFC-enabled phones that were available at that time – the Nokia 3220 with an additional NFC-shell.

It is nearly impossible to programmatically measure the time for reading an NFC-tag. First, it is difficult to detect the start of the NFC connection with the existing APIs. Second, from an interaction point of view, it is less interesting to identify the time elapsed between powering the tag and finishing the extraction of information. Rather, the time lost because of that tag reading process is of interest. The difference is that part of the reading can already be done while the user is still in the process of pointing. Therefore, in order to measure the real system response time, we designed the following experiment: First, using the Nokia 6212, a series of 50 consecutive readings were performed as fast as possible. The user was notified at each successful read with a beep and a short vibration. Second, the same person performed the same actions the same number of times but did not touch a tag and did not wait for any response. This means, that the difference of both runs leaves only the NFC reading time. Dividing this difference by the number of repetitions results in an acceptable value for system response time of 0.40 seconds for a static interface, e.g. a poster.

In addition to the direct response time between the touching of a tag and the phone's response, we repeated the experiment using a dynamic setup. This means that the notification needed to be sent via Bluetooth from the phone to the server and processed there. This resulted in a value for R_{NFC} for dynamic interfaces and the Nokia 6212 phone of 0.62 seconds. This value was also successfully validated in the video recordings from the other user studies reported later in this paper. For the previous Nokia 6131 phone, we measured R_{NFC}(Nokia 6131) to be 1.20s and for the only recently released Google Nexus S phone R_{NFC}(Nexus S, static) = 0.34s, R_{NFC}(Nexus S, dynamic) = 0.56 seconds.

4.3 Pointing / Fitts' Law Test

After the introduction to NFC and multi-tag UIs, we conducted the first part of the study to investigate pointing and selection speed. The basic setup mimicked similar studies that are used to derive the relationship between selection time and distance as well as size of the target, known as Fitts' Law [7]. The basic idea of this part of the study is to have users consecutively point to and touch specific areas. By varying the size and distance of these areas, the effect of those parameters can be measured, see [7]. We implemented both a static and a dynamic version.

Static Fitts' Law Study and Results

The static version is a direct copy of the original Fitts' Law experiment and uses three different sizes, respectively widths, of targets (1, 2 and 4 tags) and distances between them (30, 60 and 90 cm). Fig. 1 (left) shows some example configurations. In order to quickly change from one configuration to another, we created modular tiles (see Fig. 1, right) that can be easily attached and moved on a wall or whiteboard. The order of use was counterbalanced with a Latin Square design. During the study, an NFC-enabled Nokia 6212 phone was used as pointing and tag reading device. Each subject contributed with 15 measurements per setting.

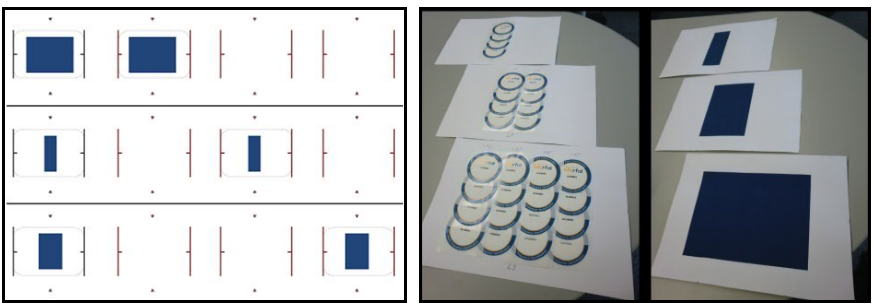


Fig. 1. Left: Three sample configurations for the Fitts' Law study with three different distances between targets (30, 60 and 90 cm). Right: Tiles for the Fitts' Law study (back side with tags and front side) with different widths (1, 2 and 4 tags).

The results of the average measurements can be seen in Table 2. Using extensive video analysis, we created two datasets, one which included runs where reading the NFC tag failed (with errors) and one that excluded those (w/o errors). Two interesting aspects can be seen: first, the effect of such errors is not visible for UI elements of width 1 tag. Second, without considering errors the times do not seem dependent on the width of the elements. Both observations can be grounded statistically: the differences between measurements including and excluding errors are not significant for a width of 1 tag ($p>0.05$) but are significant for all runs with 2 or 4 tags ($p<0.05$). Further, a linear regression analysis of the measurements with respect to Fitts' Law parameters reveals a very low correlation factor of only 0.29.

These results can be explained by the very nature of these errors. Although the NFC-tags have been attached so close to each other that they overlap, there still remains some noticeable space between two tags where the phone is not able to read any of those tags. This directly explains the first finding. The second finding can be explained similarly as the users quickly adapted to these errors and started trying to touch exactly the spot that they found out to be working. Although it should be easier to hit a wider target, users voluntarily restricted their target width to 1 tag.

Since Fitts' Law does not seem to be a good approximation, we propose to distinguish two cases: pointing to a close target (distances of about 30cm or less) from far targets. Using the overall average time for 30cm and subtracting the NFC reading time, this yields a value of P(close) = 0.58s. For other distances P = 0.83s.

Table 2. Measurements for pointing (static UI) in seconds without & with technical errors

Tile width	1 tag			2 tags			4 tags		
W/o errors	0.98s	1.05s	1.23s	0.99s	1.07s	1.26s	0.99s	1.05s	1.23s
W errors	1.00s	1.07s	1.25s	1.07s	1.12s	1.35s	1.16s	1.10s	1.34s

Dynamic Fitts' Law Study and Results

The dynamic version could have directly copied the setup from the static version. However, in order to reduce the effect of subject adaptation to errors in hitting between NFC tags, we modified it to a more dynamic version. A sequence of boxes of varying sizes appeared at varying locations on the dynamic NFC-displays, one appearing after the previous one had been touched. The sequence of positions and dimensions was randomly generated once and was the same for all subjects and all runs. Each subject was asked to do two runs, each with 21 rectangle touches, resulting in 20 measurements per run.

As the sizes and distances were random on a continuous scale, one cannot simply create a table as in the static case. When applying the necessary calculation to compute the index of difficulty (ID) of Fitts' Law from the distance and width values, the result should reveal a roughly linear relationship between ID and time. However, it turns out that the correlation factor is only 0.21. Therefore, it is difficult to justify a linear relation as postulated by Fitts' Law in the dynamic case. In line with the argumentations from the static case, we propose to apply the pointing times from the static case also in dynamic tag-based applications.

4.4 Sequences of Touch-Interactions

The second part of the study tries to find improvements to the current rules for placing Mental Act (M) operators. This is an important aspect as the number of Mental Act operators within a KLM can largely influence the overall time estimation. Set as 1.35 seconds [5], this operator is one of the largest ones. The study will shed light on when and where to place this important operator. More precisely, the effect of the following independent variables on Mental Act operator placement was measured:

1. **Sequence Length (1 / 3 / 5):** the number of tags to be read consecutively
2. **Tag Distance (close / far):** the physical distance between consecutive tags within a sequence

3. **Interaction Type (simple touching / confirmation):** presence of attention shifts between physical UI and mobile device
4. **Interface Type (static / dynamic):** the type of physical interface

We designed a set of physical UIs in order to test all parameters. The combination of the variables "Interaction" and "Interface Type" results in 4 different prototypes. First, a static and a dynamic UI have been built with the same visual setup (see Fig. 2). Second, for each UI, two mobile prototypes have been created: for the first one, touching a tag is sufficient for the tasks while for the second one, the phone displays a dialog after each touch that needs to be confirmed by pressing one of the softkeys (left or right) on the mobile phone. The mapping of the "OK" button to either of them was randomized for all tasks. This ensures that users definitely have to look at the screen. In addition, the combination of the variables "Sequence Length" and "Tag Distance" results in 6 tasks. We chose sequence lengths of 2, 4, and 6 tags (i.e. 1, 3, and 5 pointing actions from one tag to the next). Tag distance only distinguished 'close' and 'far' distances, representing distances of less and more than 40cm.

Each subject carried out each task with each prototype (counterbalanced). Again, the tasks were executed with a Nokia 6212 phone. Time measurements were done either on the phone itself (static versions) or on the server (dynamic versions). In addition, we taped all subjects in order to spot errors that occurred during the tasks.

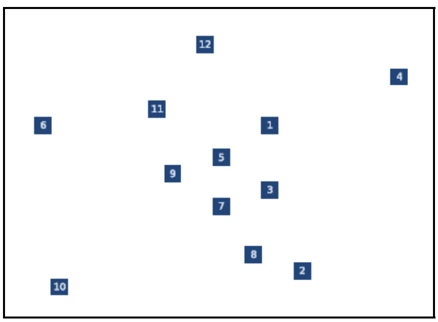


Fig. 2. NFC tag layout for the touch sequence study. Odd numbers are close together.

During the analysis of the retrieved data, all erroneous runs (user errors as well as technical errors) were discarded. The amount of errors in general was very low. Since single tag sized spots were used especially technical errors hardly ever occurred as, mentioned in the pointing experiment, gaps are the main issue for technical errors.

Interaction Type "simple touching"

This section covers the Interaction Type without forced attention shift, i.e. without needing to confirm the tag reading. Table 3 shows the average values for all values of the other 3 independent variables. It is not completely clear why there is a (in 4 of 6 cases even significant, on average 16%) slow-down of the interaction in the dynamic version compared to the static system. Some comments from the subjects indicate that a slightly more glaring and slightly less clear dynamic display in contrast to the matt and high resolution printout might have made the interaction slightly more difficult.

However, we do not have strong evidence for that and a more detailed inspection of this effect will remain future work.

Two further observations can be made: first, as expected, measured times for the Tag Distance "far" experiments are higher than for Tag Distance "close". Second, and perhaps more interesting, interaction times increase linearly when increasing the number of tags to be read. The 4 combinations far/dynamic, far/static, close/dynamic, and close/static reveal a very good linearity with a R^2 of 0.89, 0.93, 0.85, and 0.97, respectively. This strongly indicates that there is no need to place any additional Mental Act operators for longer interaction sequences. Of course, this result is limited to sequences of no more than 6 tags, which however, in our experience, seems reasonable with most interaction scenarios. Thus, we conjecture: **Sequence Length and Tag Distance do not influence M placement.**

Table 3. Total performance times for Interaction Type "simple touching"

Task	Close, 1	Close, 3	Close, 5	Far, 1	Far, 3	Far, 5
Static	1.00s	3.19s	5.31s	1.29s	3.87s	6.32s
std. dev.	0.10s	0.32s	0.38s	0.33s	0.61s	0.66s
Dynamic	1.29s	3.31s	6.58s	1.47s	4.41s	7.16s
std. dev.	0.32s	0.38s	1.39s	0.62s	0.81s	0.99s
Significance	p¡0.01	ns	p¡0.01	ns	p¡0.05	p¡0.05

Interaction Type "confirmation"

This Interaction Type uses forced attention shift to model mobile interactions with physical UIs that require users to switch their attention to the mobile device, e.g. for confirmation or input from the keyboard. Table 3 shows the averages for all values of the other 3 independent variables. The same effects as with the Interaction Type "simple touch" can be observed. The slow-down of dynamic interface interaction with respect to the static version is again measurable (significant for all cases, on average 26%).

Table 4. Total performance times for Interaction Type "confirmation"

Task	Close, 1	Close, 3	Close, 5	Far, 1	Far, 3	Far, 5
Static	1.76s	5.50s	10.58s	2.28s	7.51s	12.10s
std. dev.	0.17s	0.36s	1.11s	0.41s	0.98s	1.22s
Dynamic	2.42s	7.47s	12.89s	2.75s	8.89s	14.94s
std. dev.	0.43s	1.35s	1.32s	0.69s	1.55s	2.82s
Significance	p¡0.01	p¡0.01	p¡0.01	p¡0.05	p¡0.01	p¡0.01

Furthermore, the two previous observations can again be made: first, measured times for the Tag Distance (far) experiments are higher than for Tag Distance (close). Second, interaction times increase linearly when increasing the number of tags to be read. The 4 combinations far/dynamic, far/static, close/dynamic, and close/static also reveal a very good linearity with a R^2 of 0.88, 0.95, 0.93, and 0.96, respectively. This strongly supports the conjecture from the "simple touch" experiment.

In addition, a video analysis showed that attention shifts were in fact made, but that subjects became used to this very quickly and became able to habitually glance at the display and press the correct button, in most cases even without retracting the mobile device from the physical UI. Pauses of any kind before attention shifts were not observed which indicates that in such repetitive sequences mental preparations play a very minor role. Thus, we conjecture: **Interaction Type (attention shifts) does not influence M placement.** Also, even though there was a significant difference between dynamic and static interfaces (Interface Type), it most likely results from aspects inherent to the projection setup but definitely cannot be circumvented by adding Ms. **Interface Type (static / dynamic) does not influence M placement.**

Recommendation for Mental Act Operator Placement

In our experience, hardly any mental preparation happens during touch sequences, as the rather accurate modeling of the measured times without Ms show. It was however observed, that before a sequence subjects did actually pause for a brief time which could be assumed as mental preparation. Thus, we propose to generalize the placement rules of M and our recommendation can be summarized into: **Add Mental Act (M) operators before cognitive chunks.**

Based on their concept in cognitive psychology, Card et al. [5] define chunks as “highly integrated sub-methods that show up over and over again in different methods." These chunks “usually reflect syntactic constituents of the system's command language." Although the identification of such chunks can be difficult, this recommendation matches more with the guidelines presented by Kieras [13]. The most important advice is to identify chunks and thus place Ms consistently, especially when pitting different UI designs against each other. The assumption is, that making the same placement ‘mistakes’ in all alternative designs still yields a good comparison although the estimated times might highly deviate from actual expert execution times.

As a brief illustration of the impact of the new model proposal, take the following example as described above, touching and confirming the touch on 5 tags that are located close to each other on a static interface (i.e. Sequence Length “5”, Tag Distance “close”, Interaction Type “confirmation”, Interface Type “static”). Table 4 shows an average measured value of 10.6 seconds.

Original KLM: For each of the 5 NFC reads, the user switches attention to the display S_{macro}, points to the tag P, and waits for the read to finish R(NFC). Additionally, each time, the confirmation needs a Mental Act M, an attention shift to the phone S_{macro}, and a key press K(hotkey). This adds up to 18.6 seconds.

Proposed KLM: according to our new model, the Mental Act is not included, the pointing operation uses the P(close) operator, and the value for R(NFC) is updated. This adds up to 10.5 seconds which is much closer to the measured time of 10.6s.

4.5 An Updated Mobile KLM – Summarizing the Results

Based on the studies and their results, this section summarizes the updated Mobile KLM for NFC-based mobile applications. As described in the section about shortcomings of the original Mobile KLM, four major aspects have been investigated. We now present the results of the studies with respect to each of them:

1. System response time reading NFC tags with current mobile devices
2. Model errors caused by mismatches between tag-matrix and UI widgets
3. Pointing time for close and far objects
4. Chunking of action sequences needs to be reevaluated

Item (1), see Section 4.2: System response time will always be a parameterized operator that will have to be measured for each specific setup. However, it helps future modelers to provide current values for specific devices. First, they do not have to perform exhaustive studies themselves; second, it can help to know the different values for different devices to decide which ones should be targeted at. In combination with previous work (Nokia 3220 from [10]), we now have as value for reading an NFC tag:

R_{NFC}(Nokia 3220+NFCShell) = 2.58s

R_{NFC}(Nokia 6131) = 1.20s

R_{NFC}(Nokia 6212, static) = 0.40s, R_{NFC}(Nokia 6212, dynamic) = 0.62s

R_{NFC}(Nexus S, static) = 0.34s, R_{NFC}(Nexus S, dynamic) = 0.56s

Item (2): Although errors are usually not modeled within a KLM, it does make sense to incorporate some treatment of those if they occur even despite following the expert user assumption. Mostly due to the fact that NFC devices have not been initially designed for multi-tag UIs, tag reading accuracy depends on several factors. One of the most prominent ones is the number of tags that render a displayed UI element interactive. If more than one tag is used for, e.g. a button, small gaps appear between those tags. If users touch with their phone on one of those gaps, the probability is high that the touch attempt will be missed by the system. The KLM does not include a mechanism to specify that, e.g., every 5^{th} touch needs a higher value or an additional attempt. Therefore, we propose to model these issues by adding a small amount of time to the NFC Action operator. This models additional time that occurs with some low probability over time. As this applies mostly to multi-tag interaction and can be neglected for single-tag interfaces, the operator is zero for the single-tag case. The other values result from subtracting the measurements of all reads without technical errors from those with errors and dividing this by the number of interactions. As these are on average quite different for static and dynamic interfaces, we propose using different values for each of the action operators:

A_{NFC}(single-tag) = 0.00s

A_{NFC}(multi-tag, static) = 0.07s, A_{NFC}(multi-tag, dynamic) = 0.32s

Item (3), see section 4.3: If a fine-grained model for an interface that requires many pointing operations needs to be created, the pointing operator should be implemented using a specific model such as Fitts' Law (provided there is no indication that the interface does not adhere to it). However, in practice, this is rarely done as it complicates the generation of the model. Already in the earliest uses of the model (see [5]), it has proven to be sensible to use constant average values. However, we found that the original value of about 1 second [5] is overly high. In addition, it makes a significant difference whether two consecutive touches are close (within 30cm) or not. Thus, we propose two constant values depending on the distance between source and target of a pointing operation:

P(close) = 0.58s, P = 0.83s

Item (4), see section 4.4: The Mental Act operator is of high importance in most KLMs as it usually occurs quite often and, with 1.35 seconds, all occurrences together add up to a considerable amount of time. Our proposal for a placement rule is:

Add Mental Act (M) operators before cognitive chunks.

Although the accurate identification of such a chunk is difficult, it is clear that not the specific place but only the number of such operators counts [13]. In addition, especially when comparing different interfaces, it is more important to be consistent in operator placement than to be 100% in accordance with reality. With the studies described above, we were able to provide additional insights:

- A chunk is independent of the length of a touch sequence (if the sequence is known to the user beforehand)
- Additional mental preparations are not necessary for pointing longer distances with respect to shorter ones

5 Validation of the Updated Mobile KLM

In order to find out whether the observed results yield to a valid model and to test the superiority with respect to the existing model, we conducted two additional studies with example applications to validate the applicability and accuracy of the updated model. We recruited the same subjects from previous studies to perform the additional studies. Prior to the experiments, KLM models, i.e. the operator sequences were put together both with the existing Mobile KLM and the updated version so as to compare them later on with the actual execution times and each other. Distractions were not taken into consideration since these were not present in the lab setup. It should be noted that, when modeling a task with the original Mobile KLM, we already used the updated time value for the Nokia 6212. This is in line with the interpretation of the System Response Time operator R which is to be parameterized according to the hardware in use.

5.1 Validation Scenario 1: Munich Map Poster

The first validation used a prototype by Broll and Hausen [4] that was designed for different tasks on a map of sightseeing places in the city of Munich. Examples for such tasks include finding information about specific points of interest and getting route directions. A poster with NFC-tags was used as a static physical UI and again, a Nokia 6212 with a J2ME application was used as the input device. This application measured the times between the first touch of a tag and the last user action either by touch or key press. To avoid learning effects, the tasks of this study were counterbalanced with a Latin square design. Tasks always began with touching a “start” tag which started the measurement. Prior to the actual runs and measurements, subjects were instructed to familiarize themselves with the prototype by touching various tags and exploring the phone UI. The following tasks had to be performed:

- **Task 1** merely consisted of consecutively retrieving information from two points of interests, i.e. tags. The KLM for touching 2 POIs is 2*(P, R(NFC)) which amounts to 2.80s and 2.46s for the old and new model, respectively. The differences stems from the updated pointing time. In comparison, the average of the actual user data was 2.46s.
- **Task 2** consisted of selecting 3 nearby points of interest and having the system generate a route between them. As this task is mostly a subtask of Task 3, we refer to Task 3 for the models. The time estimations are 6.07s (old model) and 4.04s (new model) with the user data averaging to 4.75s.
- **Task 3** (see **Table 4**) is similar to Task 2 but adds one far pointing operation and increases the route size to 5 POIs. The results are summarized in Table 4 which shows the predictions for the third task as modeled and calculated with the original Mobile KLM, with our proposed updated KLM as well as the average from the performance time measured with the study. The targets of all pointing operations are quite close which is reflected in the new model. Also, the original KLM would force an M operator before a macro attention shift. The reason this is not done in the updated KLM is because it can be seen as part of a single chunk where selecting the points directly belongs to the route generating subsequence.

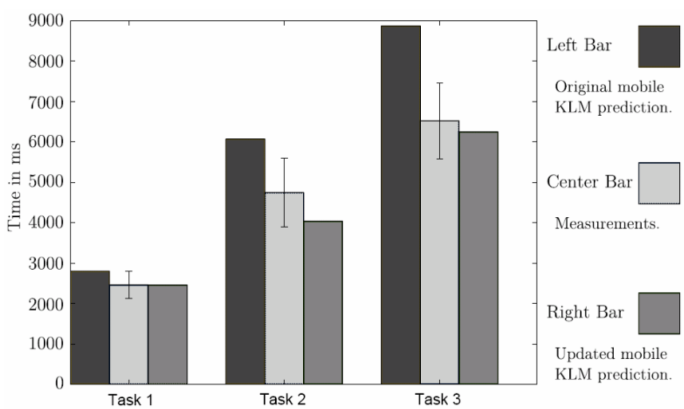


Fig. 3. Comparison of the existing model, the times measured within the study, and the new model for 3 tasks for the Munich Map Poster scenario

Fig. 3 summarizes the modeled and measured values for the 3 tasks. For the short Task 1, the original KLM predicts rather precisely but is already outperformed by the updated model. With increasing task length, the prediction error of the original Mobile KLM increases. In general, it can be observed that the original KLM tends to predict longer times than are actually measured. This is largely alleviated by the new model which tends to predict shorter times, but is considerably more accurate. This result is also more in line with the original concept of KLM which is supposed to model a lower bound for execution time [5] as dictated by the expert user assumption.

Table 4. Validation "Munich Map Poster", Task 3 long route generation

Interaction	Original KLM		Updated KLM	
Touch POI_1	P, R(NFC)	1.40s	P, R(NFC)	1.23s
Touch POI_2	P, R(NFC)	1.40s	P(close), R(NFC)	0.98s
Touch POI_3	P, R(NFC)	1.40s	P(close), R(NFC)	0.98s
Touch POI_4	P, R(NFC)	1.40s	P(close), R(NFC)	0.98s
Touch POI_5	P, R(NFC)	1.40s	P(close), R(NFC)	0.98s
Look at phone	M, S(macro)	1.71s	P(short), S(macro)	0.94s
Press hotkey	K(hotkey)	0.16s	K(hotkey)	0.16s
Total		**8.87s**		**6.25s**
Study Average	**6.52s**			

5.2 Validation Scenario 2: London Map Projection

The second validation is based on a prototype that shows a map of London on a dynamic NFC-display, similar to the one that was used for the dynamic UI experiments and the introductory game before the experiments. With this prototype, users can touch the dynamic NFC-display with their mobile devices to see details about hotels and sights. Again, a Nokia 6212 and a J2ME client were used in conjunction with a J2SE server producing the dynamic UI. Task execution times were measured via video analysis from the point right after bringing the phone close to an active area on the physical UI, i.e. right before the first NFC tag reading. The active areas on the dynamic UI were all the size of one tag, so technical errors as described above did not have to be taken into consideration at all. To avoid learning effects the tasks were counterbalanced with a Latin square design. Again, subjects were instructed to familiarize themselves with the prototype by touching various tags and exploring the phone UI prior to the actual runs and measurements. The following tasks had to be performed:

- **Task 1** mainly consisted of touching a tag to show all hotels, select one hotel, and touch a tag to add it to the favorites. The models consist only of P, R(NFC), and system response times. The difference of 6.54s (old model) and 5.53s (new model) stems from the changed time for the Pointing operator P. The measured average time was 5.39s.
- **Task 2** merely increases the length of the task. The difference between the existing and the updated model becomes more pronounced as the time estimates for the pointing operations were estimated as too high by the existing model. Results are: 11.7s (old model), 9.69s (new model), and 10.2s for the study measurements.
- **Task 3** consisted of a longer task. First, a sight needed to be added to the favorites. Second, this sight should again be removed from the favorites; however, the phone menu should be used for this instead of the tag matrix interface. The results are summarized in Table 5 which shows the predictions for the third task as modeled and calculated with the original Mobile KLM, with our proposed updated KLM as well as the average from the performance time measured with the study subjects. The application dependent constants R_0=0.20s, R_1=0.60s, R_2=0.57s, R_3=0.57s, and R_4=1.97s are System Response Time operators that we measured beforehand.

Table 5. Validation "London Map Projection", Task 3 using NFC and phone interface

Interaction	**Original KLM**		**Updated KLM**	
Touch tag (show sights)	R(NFC), R_1	1.00s	R(NFC), R_1	1.00s
Touch POI	P, R(NFC), R_2	1.97s	P, R(NFC), R_2	1.80s
Touch tag (add to)	P, R(NFC), R_0	1.60s	P(close), R(NFC), R_0	1.18s
Touch POI	P, R(NFC), R_0	1.60s	P, R(NFC), R_0	1.18s
Look at phone	M, S(macro)	1.71s	P(short), S(macro)	0.94s
Navigate to favorites	3*(M, K(hotkey))	4.53s	M, 3*K(hotkey)	1.83s
Open favorites	K(hotkey)	0.16s	K(hotkey)	0.16s
Select POI	K(hotkey), R_4	2.13s	K(hotkey), R_4	2.13s
Remove POI	M, K(hotkey), R_3	2.08s	K(hotkey), R_3	0.73s
Total		**16.8s**		**11.0s**
Study Average	**13.2s**			

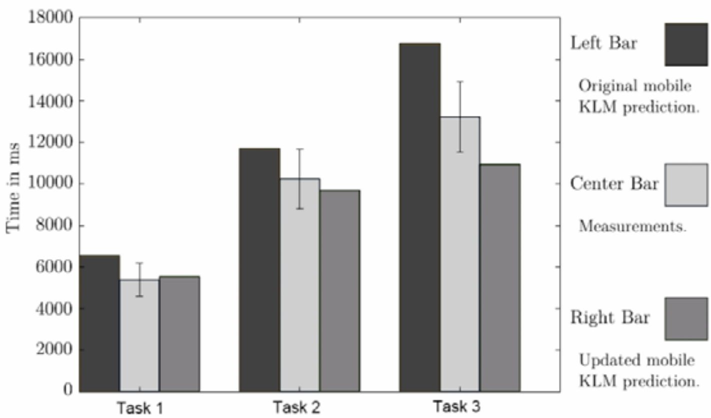


Fig. 4. Comparison of the existing model, the times measured within the study, and the new model for 3 tasks for the London Map Projection scenario

For the models, we had to measure some system response times that occur when waiting for all hotels being displayed, show information on the phone, etc. These were the same for the old and the new model. Besides the reduced time for pointing operations, the largest difference stems from the difference in M placement. In contrast to the new model, the original Mobile KLM prediction forces an M in front of the Macro Attention Shift for looking at the phone UI. Also, when navigating the mobile menu, the original Mobile KLM would place an M before nearly each key press [10]. Fig.4 summarizes the modeled and measured values for the 3 tasks.

Though deviating, the original Mobile KLM seems to predict execution times quite accurately for Tasks 1 and 2. However, the prediction for task C is way off. It predicts a 27% longer time than the measurements show. Except for one case (the predicted time for Task 1 is about 3% longer than the measured time), the times predicted by the updated KLM are shorter than the measured times which is a good indication for the intended applicability of the KLM for predicting a lower bound for execution time [5]. In summary, the two validation experiments very well illustrated the increased accuracy of the four aspects that have been identified for a need to update.

6 Summary and Conclusion

The Keystroke-Level Model (KLM, [5]) has already been used to model interactions and predict task execution times for a great variety of applications. For advanced interaction with mobile devices, an extension to the original, desktop-based KLM has been described in [10] (Mobile KLM). In this paper, we have revisited and revised the Mobile KLM in several aspects, especially focusing on interactions with interfaces featuring multiple NFC tags. These improvements have been made taking two important categories of such (multiply) tagged applications into account: static and dynamic interfaces. Main insights generated by our work include:

- Updated KLM values for the system response time of current phones for detecting and reading NFC tags
- Explanation and guideline to avoid errors using multiply tagged interfaces (use odd number of tags) and a way to model errors for those instances
- A more precise estimation of pointing time (distinguishing between ‘close’ and ‘far’ pointing, no difference between static and dynamic interfaces)
- Improved guidelines for placing Mental Act (M) operators

With these additions, it is possible for designers and developers to model interfaces that make use of advanced, multi-tagged UIs with the easy-to-use KLM. Several metrics for user interfaces such as user acceptance or fun are of course difficult to evaluate with user models. Still, the revised Mobile KLM allows a quick and simple estimation of end-user task performance times as well as a concise and meaningful comparison between interfaces. The main advantage of this approach is that such metrics can be generated quickly, cheaply, and without the prior implementation of any of the interfaces.

We thus hope that our work broadens the applicability of the KLM to novel and upcoming applications and improves the accuracy of the approach while keeping the simplicity of the model intact. Of course, this type of work will be ongoing as long as new types of interactions will be added. In general, including our previous and related work, we have thus far managed to cover a very broad array of interactions with mobile devices.

References

1. Bälter, O.: Keystroke Level Analysis of Email Message Organization. In: Proc. CHI 2000, pp. 105–112 (2000)
2. Broll, G., Siorpaes, S., Rukzio, E., Paolucci, M., Hamard, J., Wagner, M., Schmidt, A.: Supporting Mobile Service Usage through Physical Mobile Interaction. In: Proc. PerCom 2007, pp. 262–271 (2007)
3. Broll, G., Graebsch, R., Holleis, P., Wagner, M.: Touch to Play - Mobile Gaming with Dynamic, NFC-based Physical User Interfaces. In: Proc. MobileHCI 2010 (2010)

4. Broll, G., Hausen, D.: Mobile and Physical User Interfaces for NFC-based Mobile Interaction with Multiple Tags. In: Proc. MobileHCI 2010 (2010)
5. Card, S.K., Moran, T.P., Newell, A.: The Keystroke-Level Model for User Performance Time with Interactive Systems. Communications of the ACM 23(7), 396–410 (1980)
6. Card, S.K., Newell, A., Moran, T.P.: The Psychology of Human-Computer Interaction. Lawrence Erlbaum Associates, Inc., Mahwah (1983)
7. Fitts, P.M.: The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement. Journal of Experimental Psychology 47(6), 381–391 (1954)
8. Hardy, R., Rukzio, E.: Touch&Interact: Touch-based Interaction of Mobile Phones with Displays. In: Proc. MobileHCI 2008, pp. 245–254 (2008)
9. Haunold, P., Kuhn, W.: A Keystroke Level Analysis of a Graphics Application: Manual Map Digitizing. In: Proc. CHI 1994, pp. 337–343 (1994)
10. Holleis, P., Otto, F., Hussmann, H., Schmidt, A.: Keystroke-level Model for Advanced Mobile Phone Interaction. In: Proc. CHI 2007, pp. 1505–1514 (2007)
11. Holleis, P.: Integrating Usability Models into Pervasive Application Development, PhD Thesis, Ludwig-Maximilians-Universität München, Germany (2009)
12. John, B.E., Kieras, D.E.: The GOMS Family of Analysis Techniques: Tools for Design and Evaluation. Human-Computer Interaction Institute. Technical Report (1994)
13. Kieras, D.: Using the Keystroke-Level Model to Estimate Execution Times. The University of Michigan. Technical Report (1993), http://www.pitt.edu/~cmlewis/KSM.pdf
14. Koester, H.H., Levine, S.P.: Validation of a Keystroke-Level Model for a Text Entry System Used by People with Disabilities. In: Proc. Assets 1994, pp. 115–122 (1994)
15. Luo, L., Siewiorek, D.P.: KLEM: A Method for Predicting User Interaction Time and System Energy Consumption during Application Design. In: Aberer, K., Choi, K.-S., Noy, N., Allemang, D., Lee, K.-I., Nixon, L.J.B., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., Cudré-Mauroux, P. (eds.) ASWC 2007 and ISWC 2007. LNCS, vol. 4825, Springer, Heidelberg (2007)
16. Manes, D., Green, P., Hunter, D.: Prediction of Destination Entry and Retrieval Times Using Keystroke-Level Models. Technical Report. University of Michigan (2007)
17. Mori, R., Matsunobe, T., Yamaoka, T.: A Task Operation Prediction Time Computation Based on GOMS-KLM Improved for the Cellular Phone and the Verification of that Validity. In: Proc. ADC 2003 (2003)
18. Ramírez-González, G., Muñoz-Organero, M., Kloos, C.D., Astaiza, Á.C.: Exploring NFC Interactive Panel. In: Proc. Mobiquitous 2008 (2008)
19. Reilly, D., Welsman-Dinelle, M., Bate, C., Inkpen, K.: Just Point and Click?: Using Handhelds to Interact with Paper Maps. In: Proc. MobileHCI 2005, pp. 239–242 (2005)
20. Riekki, J., Salminen, T., Alakärppä, I.: Requesting Pervasive Services by Touching RFID Tags. IEEE Pervasive Computing 5(1), 40–46 (2006)
21. Rohs, M., Gfeller, B.: Using Camera-Equipped Mobile Phones for Interacting with Real-World Objects. In: Proc. Advances in Pervasive Computing, pp. 265–271 (2004)
22. Sánchez, I., Riekki, J., Pyykknen, M.: Touch&Control: Interacting with Services by Touching RFID Tags. In: Proc. IWRT 2008 (2008)
23. Seewoonauth, K., Rukzio, E., Hardy, R., Holleis, P.: Touch&Connect and Touch&Select: Interacting with a Computer by Touching it with a Mobile Phone. In: Proc. MobileHCI 2009, pp. 1–9 (2009)
24. Vetter, J., Hamard, J., Paolucci, M., Rukzio, E., Schmidt, A.: Physical Mobile Interaction with Dynamic Physical Objects. In: Proc. MobileHCI 2007 (2007)
25. Want, R.: An Introduction to RFID Technology. IEEE Pervasive Computing 5(1), 25 (2006)

The Entropy of a Rapid Aimed Movement: Fitts' Index of Difficulty versus Shannon's Entropy

R. William Soukoreff[1], Jian Zhao[1], and Xiangshi Ren[2]

[1] Department of Computer Science, University of Toronto, Canada
[2] The School of Information, Kochi University of Technology, Japan
{will,jianzhao}@dgp.toronto.edu,
ren.xiangshi@kochi-tech.ac.jp

Abstract. A thought experiment is proposed that reveals a difference between Fitts' index of difficulty and Shannon's entropy, in the quantification of the information content of a series of rapid aimed movements. This implies that the contemporary Shannon formulation of the index of difficulty is similar to, but not identical to, entropy. Preliminary work is reported toward developing a model that resolves the problem. Starting from first principles (information theory), a formulation for the entropy of a Fitts' law style rapid aimed movement is derived, that is similar in form to the traditional formulation. Empirical data from Fitts' 1954 paper are analysed, demonstrating that the new model fits empirical data as well as the current standard approach. The novel formulation is promising because it accurately describes human movement data, while also being derived from first principles (using information theory), thus providing insight into the underlying cause of Fitts' law.

Keywords: Fitts' law, Human Performance Modelling, Entropy.

1 Introduction

Fitts' law [3] describes the relationship between movement time, distance, and target width, for people engaged in rapid aimed movements. See Figure 1.

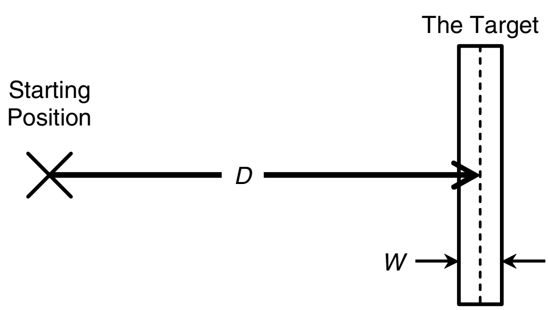


Fig. 1. The Fitts' law movement paradigm

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 222–239, 2011.

The *index of difficulty* of a given movement is defined,

$$ID = \log_2\left(\frac{D}{W}+1\right) \quad (1)$$

where D represents the distance from the origin of the movement to the centre of the target, and W denotes the width of the target, both typically measured in millimetres or pixels. Equation 1 is the *Shannon formulation* of the index of difficulty, a formulation that is theoretically and empirically preferable to Fitts' original [8, 9].

Intuitively the index of difficulty, ID, represents the difficulty of completing the movement task defined by D and W, as ID increases for longer distances or narrower targets. By convention the units of the index of difficulty are taken to be *bits*. This convention diverges from the usual definition of bits as defined by information theory [15, 16], as the weighted average of the base-2 logarithm of the probabilities of the set of outcomes of a stochastic process – a point we will return to shortly.

The crux of Fitts' law is that there is a linear relationship between index of difficulty and *movement time*, the average time required to complete the movement task, typically measured in seconds or milliseconds. Commonly in the academic literature this linear relationship is written,

$$MT = a + b \times ID \quad (2)$$

where a and b are empirically determined constants, although in Fitts' original publication [3] the intercept term, a, was zero.

Fitts' index of difficulty is taken (if not formally then at least in name) to be a measure of information (viz., *bits* in the sense of entropy); a consequence of this is that the reciprocal of the slope coefficient b in Equation 2 may be interpreted as *throughput*, the rate of transmission of information, having the units *bits per second*,

$$Throughput = {}^{1}\!/_{b} = \frac{ID}{MT - a}. \quad (3)$$

Note that the ISO standard on pointing device evaluation [5] explicitly defines throughput as the ratio of the effective index of difficulty to movement time (essentially, Equation 3 with a = 0). (The technical distinction between ${}^{1}\!/_{b}$, and the ISO definition of throughput is not relevant here, under normal circumstances they are very similar, and in either case, ID is taken to be a measure of information entropy.)

Fitts' law typically yields very good fits to empirical data (with R^2 values above 0.9), and has been demonstrated to hold over a wide variety of movement conditions and circumstances (see [12], pages 281-282). In the field of human-computer interaction, Fitts' law is used to model the performance of a number of popular pointing devices including the mouse, stylus, and dactylic touch-screen – the primary applications being the prediction of movement times, and the comparative evaluation of pointing devices. When Fitts' law is used to make predictions of movement time, the meaning of Fitts' index of difficulty is of little importance, because due to the robust linear relationship between ID and MT, Fitts' law provides accurate predictions regardless of the precise meaning of the slope parameter b. However, the meaning of Fitts' index of difficulty is central to the evaluation of pointing devices, because in this case, the Fitts' throughput is used as a comparative dependent measure. This is not merely a point of theoretical interest – it is critical to the development, improvement, evaluation, and study of pointing devices that the dependent measure used to compare them is sound. It is impossible to determine the tallest man in a room solely by comparing their weights (although we might get close); similarly, it is impossible to be certain that one pointing device is superior to another if we do not know what units of measurement have been used to compare them. Without Fitts' law's throughput we would have no way to compare the over-all performance of two pointing devices, yet if Fitts' law is to be used as a performance metric, then we must know for certain that the Fitts' law throughput means what we think it means.

In the field of information theory, throughput refers to the number of useful bits of information conveyed in a period of time, or more precisely, the entropy per unit of time. Shannon [15, 16] defines entropy as a quantity of information, measured in the units, *bits*. A coin-toss (resulting in a "heads" or "tails" with equal likelihood) produces 1 bit of entropy. But this definition of entropy and throughput is very different from the terms *index of difficulty* and *throughput* as used in the context of Fitts' law. The questions we will look at next are: What is the precise relationship between the index of difficulty and entropy? Is it valid to interpret *ID* as a measure of information? And by extension, is it valid to interpret Equation 3 as throughput?

2 The Information Theoretic Basis for Fitts' Law

From the very beginning Fitts attributed his law to information theory. The hypothesis Fitts stated in his seminal 1954 paper was: "If the amplitude and tolerance limits of a task are controlled by the experimenter, and the subject is instructed to work at his maximum rate, then the average time per response will be directly proportional to the minimum average amount of *information per response*[1] demanded by the particular conditions of amplitude and tolerance" [3, page 383]. Additionally, because in his original experiments the results of using a 1 oz and 1 lb stylus were similar, Fitts concluded that the relation he discovered was due to information processing instead of a physical mechanism, "The finding that relatively small differences in performance result from the change in stylus weight, and the validity of predictions of performance rate from the index of task difficulty lend support to the basic thesis of this paper, that it is the degree of control required over the organization of a response, i.e., *the amount of information required to specify a response*, which is the major factor limiting the rate of motor performance" [3, page 390].

Crossman had a significant impact upon our understanding and treatment of Fitts' law through his invention of *the adjustment for accuracy* (a post-hoc analysis technique frequently employed in Fitts' law studies). Crossman states that entropy underpins Fitts' law: "the perceptual load in the current control of hand-movement is measured by the *difference between the initial and final entropy* of the extent of the movement" [2, page 74].

MacKenzie's contribution to Fitts' law was the Shannon formulation of the index of difficulty (Equation 1), and he states, "in executing a movement task, *ID* is the number of bits of information transmitted, and [$^1/_b$] is the rate of transmission" [9, page 98].

---

[1] Italics have been added to the quotations appearing in this section for emphasis.

In addition to the statements by the eminent contributors to Fitts' law quoted above, similar statements appear in textbooks and review papers. For example when Card, Moran & Newell discuss the comparison of pointing devices, they conclude that "the match of the Fitts' law slope of the mouse to the $[b] \approx .100$ sec/bit constant observed in other hand movement and manual control studies means that *positioning time is apparently limited by central information-processing capacities* of the eye-hand guidance system" [1, page 247]. Also Schmidt & Lee, although employing Fitts' original formulation of the index of difficulty,

$$ID_{Fitts} = \log_2\left(\frac{2D}{W}\right) \tag{4}$$

unequivocally characterise it as a quantity of information: "Recall the discussion of Hick's law of choice reaction time … the equation for that relationship also had a Log_2 term. The $\text{Log}_2(N)$, in which N was the number of equally likely stimulus-response alternatives, was a measure of the amount of information (in bits) required to resolve the uncertainty about N alternatives. The Log_2 term in Fitts' law can be seen in a similar way: $2D/W$ is related to the number of possible movements, and the $\text{Log}_2(2D/W)$ is the *information required (in bits)* to resolve the uncertainty among them" [13, page 212]. Lastly, in a recent review paper Seow writes: "This index of performance [$^1/_b$], or commonly called throughput (*TP*) by some HCI researchers … is measured in bits per unit time and is homologous to the rate of gain of information in Hick's (1952) paradigm and analogous to the channel capacity in Shannon and Weaver's (1949) theory" [14, page 332].

These quotations support the perspective that the index of difficulty, *ID*, is a measure of information, and that the reciprocal of the slope, $^1/_b$, is throughput; however, this perspective is not universal. Both Kvålseth [6] and Lai et al. [7] provide analyses of empirical data indicating that the original formulation of the index of difficulty (Equation 4) differs from entropy. However, the divergence reported by Kvålseth could arise from his participants overemphasising speed at the expense of accuracy, in the absence of the adjustment for accuracy. Lai et al. analysed only the entropy of the spread of movement endpoints, as opposed to the entropy of the entire movement tasks, which undermines their argument.

The conclusion that one may draw from the literature is that the index of difficulty is possibly (even perhaps, likely) a measure of information, but it may not be. Next we shall consider whether our doubts regarding the entropic nature of the index of difficulty can be resolved by examining the derivation and foundation of the index of difficulty.

2.1 Shannon's Theorem 17

Fitts cited Shannon's theorem 17 as the basis of his law [3, page 382]. Shannon's theorem 17 [16, page 43] describes the information capacity of a band limited continuous channel under an average power limitation in the presence of noise, finding that the channel capacity is equal to the difference between the entropy of the signal and the noise, or,

$$C = H(signal) - H(noise) = B \cdot \log_2 \frac{S+N}{N} \quad (5)$$

where $H(\cdot)$ represents the entropy of the given quantity, B represents the bandwidth of the communications channel, and S and N represent the average power of the signal and noise respectively. Fitts' original formulation of the index of difficulty (Equation 4) has been largely supplanted in the recent literature by the Shannon formulation (Equation 1), originally proposed by MacKenzie [8, 9]. Comparing Equations 1 and 5, the analogy is clear to see. Fitts took movement distance to be analogous to signal power, and target width to noise power; so the theoretical foundation that Fitts' law rests upon is an analogy to information theory. This is not a robust foundation – movement distance and target width are simply not equivalent to average power levels, which are the product of root-mean-squared voltage and current levels. These equations look similar to one another, but similarity is not enough to convince us one way or the other. Fitts' law agrees with human performance data very well under a wide variety of circumstances, suggesting that the analogy to information theory is a fitting one, but in what subtle way may it differ? Over the years several researchers have proposed alternate derivations (or formulations) of Fitts' law, usually rooted in the characteristics of movement biophysics, motor control theory, or feedback loops (see [11, 12]), however currently no universally accepted explanation of Fitts' law exists (see [19] page 761, and footnote 17).

Without a robust theoretical basis underpinning Fitts' law, not only are we unable to answer the index of difficulty versus entropy question, but we also lack an explanation for why the Fitts' law phenomena occurs – why is there a linear relationship between movement time and the index of difficulty?

3 Index of Difficulty versus Entropy – A Thought Experiment

An inconsistency between Fitts' index of difficulty and Shannon's entropy can be demonstrated via a thought-experiment that compares how these two quantities measure information transmission during an imaginary data entry task. The idea is to compare the information transmitted during a data entry single-finger typing task, as predicted by the index of difficulty of the finger movements versus the entropy of the typed text. See Figure 2.

The idea that Fitts' law could be used to predict data entry times, and consequently could be used to improve keyboard designs, originated with Card, Moran & Newell [1, pages 51-57], but was extended and popularised by others (for a review see [10]). The idea is depicted in Figure 3 – Fitts' law can be used to model the time to move between keys on a keypad or keyboard. By applying Fitts' law to the entirety of the keyboard (i.e., from every key to every key), and coupling this with a statistical model of the pairs of letters in a language, a model of single-finger text entry can be constructed.

Empirically, the model described above applies best to highly-trained "expert" data entry behaviour, where typing-time consists solely of movement time, no time being needed for searching for the next key, cognitive processing, fatigue, etc. However, in our thought-experiment, the expertise of the hypothetical data entry clerk has no effect upon the outcome because the results of this thought-experiment concern quantifying the amount of information transmitted *per keystroke*, not *per unit of time*. Thus the implications are independent of the speed of typing during the imaginary data entry task.

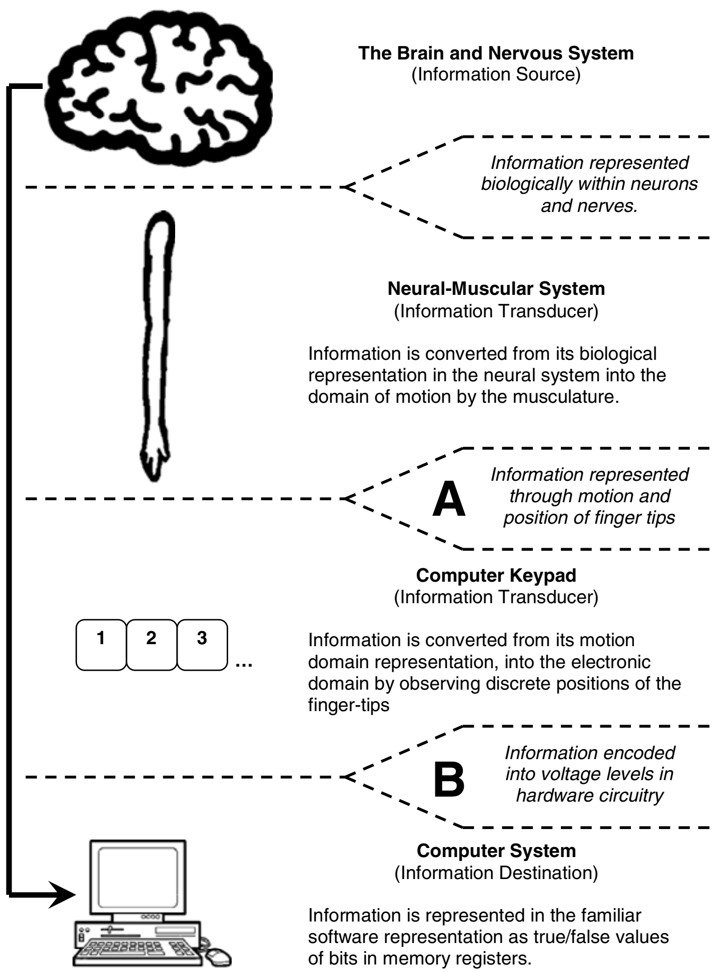


Fig. 2. The task of typing digital numeric characters is represented as a communications network containing an information source and destination with two intervening information transducers. Two specific locations within the network are identified. The point *A* corresponds to the interface between the *motion* transmission medium and the transmission medium of electronics within the keypad – the information content of the data stream at point *A* can be measured with the index of difficulty. The point *B* denotes the place within the computer where the information is represented as a series of characters – and entropy can be measured using Shannon's definition of the entropy of discrete data.

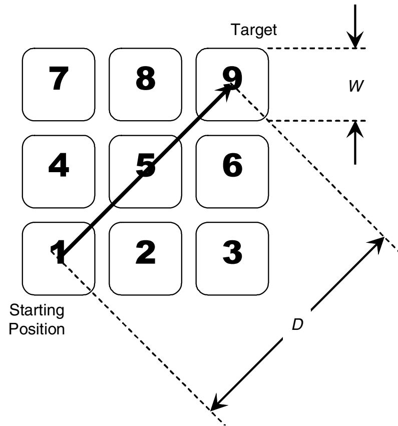


Fig. 3. Single-finger typing is depicted as a Fitts movement task. The distance between the keys, and the size of the keys[2], can be found from the geometry of the keypad. Application of Equation 1 yields the index of difficulty that corresponds to any key-to-key movement made by the finger across the keypad, representing the quantity of information needed to accomplish the keystroke according to Fitts' law. By averaging over all such movements the mean index of difficulty per character of data can be found, and if the index of difficulty is indeed a measure of entropy, then this quantity should equal the average entropy per character.

The apparatus for our thought-experiment is the *two-key binary keyboard* – an imaginary keyboard that contains only two keys, labelled '0' and '1'. A degenerate keyboard such as this would not facilitate as easy text entry as a full-size Qwerty keyboard, and yet it would be universal – with a suitable decimal, ASCII, or unicode-to-binary conversion table (and some patience), any text or numeric data could be entered. A movement model has been constructed for the two-key binary keyboard, see Figure 4.

Figure 4 presents the index of difficulty corresponding to all four possible key-to-key transition-movements (i.e., moving from a previous key to the subsequent key), that could occur during data entry on the two-key binary keyboard: 0→0, 0→1, 1→0, and 1→1. (Note that "0→0" and "1→1" denote key repetition of the '0' and '1' key respectively. The concern that the reader may have with regard to using Fitts' law to model key repeats will be dispelled in section 3.2.)

By coupling this movement model with a statistical language model it is possible to determine the average number of bits per character predicted by the Fitts' law movement model. Imagine that the "language" to be entered on the two-key binary keyboard consists of a random sequence of binary digits ('0' and '1'), where both digits have the same likelihood of occurrence,

$$p(0) = p(1) = \tfrac{1}{2} , \qquad (6)$$

[2] Key width varies with the angle of approach to rectangular keys, this is not a factor in the 1-dimensional thought experiment described in the text, so we disregard this detail here.

and occur independently of one other,

$$p(0|0) = p(0|1) = p(1|0) = p(1|1) = \tfrac{1}{2} \cdot$$

In this case the likelihood of each of the four key-to-key transitions is ¼, because,

$$\begin{aligned} p(0 \to 0) &= p(0) \cdot p(0|0) = \tfrac{1}{2} \cdot \tfrac{1}{2} = \tfrac{1}{4}, \\ p(0 \to 1) &= p(0) \cdot p(1|0) = \tfrac{1}{2} \cdot \tfrac{1}{2} = \tfrac{1}{4}, \\ p(1 \to 0) &= p(1) \cdot p(0|1) = \tfrac{1}{2} \cdot \tfrac{1}{2} = \tfrac{1}{4}, \text{ and} \\ p(1 \to 1) &= p(1) \cdot p(1|1) = \tfrac{1}{2} \cdot \tfrac{1}{2} = \tfrac{1}{4} \cdot \end{aligned} \tag{7}$$

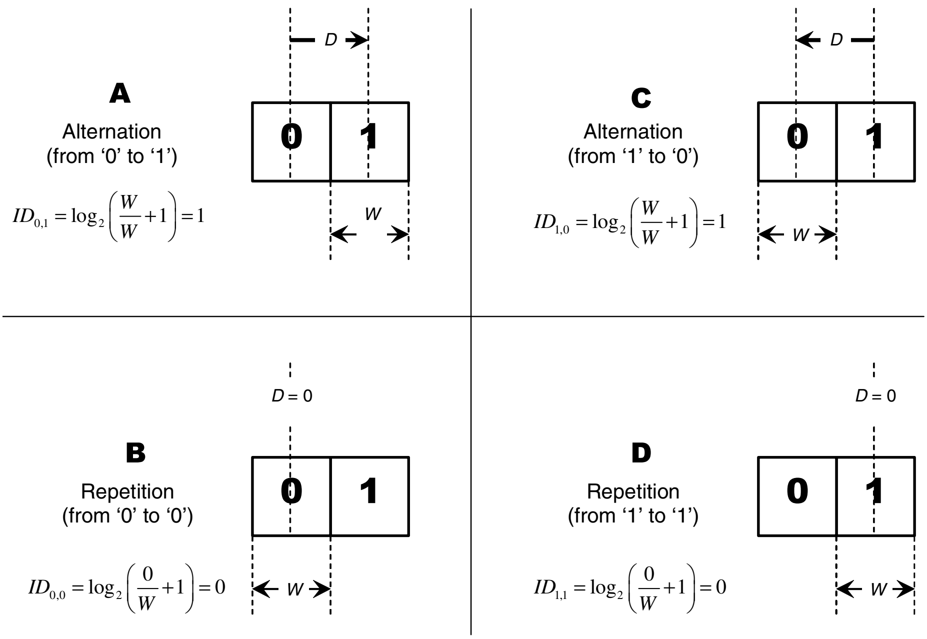


Fig. 4. The movement model for the two-key binary keyboard consists of all of the possible movements that may occur during data entry. Four possible key-to-key transitions are shown, 0→0, 0→1, 1→0, and 1→1.

3.1 Analysis of the Thought Experiment

The average number of Fitts’ law bits per key-to-key transition, is simply the average of the index of difficulties weighted by the likelihood of each key-to-key transition. So, calculating a weighted average of the index of difficulty values from Figure 4 using the transition probabilities from Equation 7 we obtain,

$$\begin{aligned}\sum_{i,j\in\{0,1\}} p(\mathrm{i}\rightarrow \mathrm{j})\cdot ID_{i,j} &= p(0\rightarrow 0)\cdot ID_{0,0} + p(0\rightarrow 1)\cdot ID_{0,1} \\ &\qquad + p(1\rightarrow 0)\cdot ID_{1,0} + p(1\rightarrow 1)\cdot ID_{1,1} \\ &= \tfrac{1}{4}\cdot 0 + \tfrac{1}{4}\cdot 1 + \tfrac{1}{4}\cdot 1 + \tfrac{1}{4}\cdot 0 \\ &= \tfrac{1}{2} \text{ bit per key-to-key transition.}\end{aligned} \tag{8}$$

According to Fitts' law, given the language model consisting of random binary digits, the average number of bits per key-to-key transition on the two-key binary keyboard is 0.5 bits. And because there is a one-to-one correspondence between keystrokes and key-to-key transitions (each keystroke being preceded by a Fitts key-to-key transition)[3], the act of entering a random sequence of binary digits results in the transmission of 0.5 bits of "Fitts' information" per keystroke. However, this quantity does not match the number of bits of information transmitted as defined by Shannon [16] Theorem 2. Shannon defines entropy (information content arising from the outcome of a stochastic process) as,

$$H(\{p_i\}) = -\sum p_i \cdot \log_2(p_i) \tag{9}$$

Substituting the probabilities defined in Equation 6 into Equation 9 yields,

$$\begin{aligned}H(\{\text{binary digit}\}) &= -\sum p_i \cdot \log_2(p_i) \\ &= -p(0)\cdot \log_2 p(0) - p(1)\cdot \log_2 p(1) \\ &= -\tfrac{1}{2}\cdot \log_2(\tfrac{1}{2}) - \tfrac{1}{2}\cdot \log_2(\tfrac{1}{2}) \\ &= 1 \text{ bit per binary digit.}\end{aligned} \tag{10}$$

Comparing the results in Equations 8 and 10 it is clear that Fitts' law and Shannon's entropy assign different quantities of information to the same task. This disparity is very troubling in a theoretical sense, because it implies that a data entry clerk can perform a series of physical motions averaging only 0.5 bits of information each (at point *A* in Figure 2), yet somehow when these motions are interpreted by the two-key binary keyboard, one entire bit of information is communicated per movement (at point *B* in Figure 2). Because there is only one communications channel between the information source and the destination, and hence no other means for the missing ½ bit of information to be transmitted to the keyboard, there appears to be spontaneous generation of information, contradicting a fundamental tenet of information theory (specifically, Theorem 7 on page 15 of Shannon [16], which states that a finite-state transducer cannot increase the entropy of a signal). It should not be possible for there to be less information at point *A* than at point *B* of Figure 2.

In this analysis Shannon's quantity of entropy was calculated via the application of a very straightforward equation (Equation 10), and the result matches our intuition – we would expect each digit in a sequence of (equiprobable and independent) binary

---

[3] We assume that the clerk positions their finger over the keyboard prior to commencing the data entry task, so that even the first keystroke is preceded by a key-to-key transition.

digits to contain one bit of information – and so it seems logical to conclude that the problem lies in the Fitts' law side of the analysis. The index of difficulty does not account for all of the information transmitted via motion.

3.2 Key Repetitions Are Not the Problem

Looking at Figures 4B and 4D, it seems that the cause of the Fitts-Shannon disparity may be the key repetition keystrokes. Key repetition is a degenerate form of "movement" where the distance moved is zero, and the corresponding Fitts' law index of difficulty for key repeats (defined by Equation 1) is also zero. In the case of the two-key binary keyboard, key repeats account for half of all key-to-key transitions, and consequently in our thought-experiment, key repetition plays a significant role in reducing the average index of difficulty. Further there is a theoretical consideration, it could be argued that Fitts' law applies only to rapid aimed *movements* – it is not entirely clear whether it can or should be used to model in-place tapping (this argument has been made before, see [18]).

We can investigate the key repetition question in two ways. (1) By enlarging the keyboard used in our thought-experiment we can reduce the frequency of key repeats. For example, in the case of a three-key keyboard, key repetition represents only 33% of the keystrokes, and for a four-key keyboard only 25%. And in general, an $n \times 1$ key keyboard would have n^2 possible unique key-to-key transitions, of which only n would be key repeats, and so the effect upon our model of key repetition can be reduced to insignificance by simply choosing a large enough value for n.[4] (2) Alternatively, we can repeat the same thought-experiment calculations with a modified language model that omits key repeats (i.e., the sequence of digits to be entered by the hypothetical data entry clerk is perfectly random and equiprobable, except that it contains no key repeats). Figure 5 depicts the results of analysing theoretical keyboards with up to $n = 30$ keys, in a similar manner to the thought-experiment described previously; the keys of the keyboards were presumed to be in an $n \times 1$ physical arrangement, and in each case the alphabet size was presumed to be n. This calculation is presented twice – once including key repeats (see the *with key repeats* curve), and once where key repeats were excluded (*without key repeats* curve).

Mathematically, the formulation for the *with key repeats* curve presented in Figure 5 is a parametric combination of two formulae that both take n (the number of keys on the keyboard) as a parameter. The ordinate (y-coordinate) side of the parametric pairing is the Fitts' index of difficulty, obtained by generalising the average index of difficulty calculation from Equations 1 and 8,

$$\overline{ID} = \frac{1}{n^2} \cdot \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \log_2 \left(|i-j| + 1 \right), \tag{11}$$

[4] Note that the language model must correspond to the size of the keyboard in use, so the data to be entered should be random ternary digits when $n = 3$, quaternary digits when $n = 4$, and in general, n-ary for larger values of n.

which corresponds to the weighted average of a series of index of difficulties (Equation 1), with target width $W = 1$, and movement distances, $D = |i - j|$, that correspond to the distance between keys i and j. The abscissa (x-coordinate) side of the parametric pairing, the Shannon side of the relationship, was generalized from Equation 9, presuming an alphabet of n independent and equally probable symbols, where $p = \frac{1}{n}$,

$$\begin{aligned} H(n) &= -\sum p_i \cdot \log_2(p_i) \\ &= \sum_{i=1}^{n} \frac{1}{n} \cdot \log_2(n) \\ &= \log_2(n). \end{aligned} \tag{12}$$

Note the implicit assumption in the definition of entropy given in Equation 12, that the data entry activity is performed without errors. As already described, our interest is in the rate of information per keystroke, not per unit of time, and so the data entry clerk in the thought experiment is free to type as slowly as they wish, and consequently may reduce the error rate to a negligible level. This formulation of entropy is valid given the assumption of negligible errors.

The formulation of the *without key repeats* curve presented in Figure 5 is also a parametric combination of two formulae, derived by removing the repeat conditions from Equations 11 and 12. The ordinate was calculated via,

$$\overline{ID} = \frac{1}{n^2 - n} \cdot \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \log_2(|i - j| + 1), \tag{13}$$

while the abscissa was,

$$H(n) = \log_2(n - 1). \tag{14}$$

The results depicted in Figure 5 demonstrate that the divergence between the Fitts' index of difficulty and Shannon's entropy are systemic, and not the result of key repeats. The trend of the divergence between the Fitts and Shannon models is clear – as the size of the keyboard increases the models do not converge (neither curve, *with key repeats* nor *without*, converge toward the reference line), a non-vanishing difference between these curves and the reference line is apparent. As we would expect, the *with key repeats* and *without key repeats* curves converge towards one another as n increases, as the effect of key repeats diminishes with increasing n. The data points toward the right of Figure 5 on the *with key repeats* curve represent the situation where the keyboard contains 30 keys, the alphabet contains 30 unique characters, and key repetition accounts for merely 3% of the keystrokes. At this point, the difference between the average entropy (4.91 bits) and the average index of difficulty (3.07 bits) is almost 2 bits, representing a significant disparity between the index of difficulty and entropy.

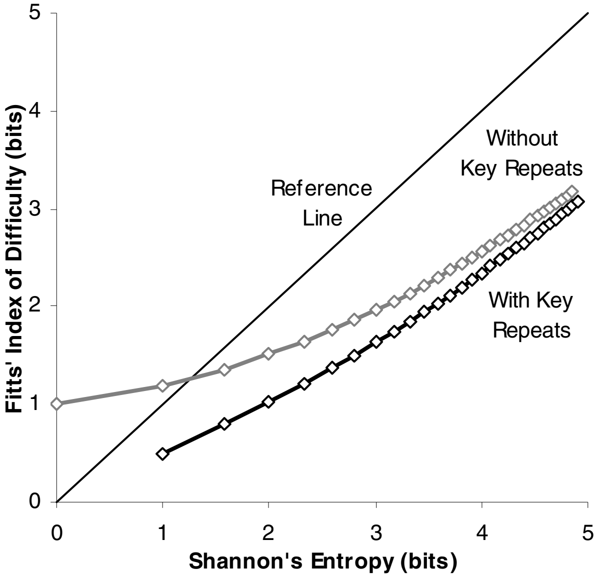


Fig. 5. Thought-experiment results for keyboards with 2 to 30 keys, with and without key repeats. The *Reference Line* (through the origin, with a slope of one) marks the position where the data points should lay if there was equality between the Fitts and Shannon models.

3.3 The Validity of the Comparison within the Thought Experiment

Could the disparity revealed by the thought experiment be an indication that the comparison between the index of difficulty and entropy proposed by the thought experiment is invalid – like "comparing apples and oranges"? No. The thought experiment constructs a series of Fitts movements for which it is possible to know the precise real entropy. This means that the thought experiment provides a means to quantify from outside of the body (entropy), the minimum information internal to the body responsible for the movement of the fingertip (Fitts' law). Or in other words, the thought experiment allows us to compare the quantity of information *encoded within* the movement, with the information necessary to *prescribe* the motion.

As the data path depicted in Figure 2 is the sole means of communication from the information source to the destination, the information must traverse the point A, wherein the data stream is entirely encoded in the movements of the operator's fingertip. The fingertip movements are interpreted by the keypad as it converts the movement signals into digits, and because the information is conveyed via the movement of the fingertip, it is the motions themselves that must contain an adequate quantity of entropy for this communication to be possible. So, the movements of the fingertip can be equivalently described as either a series of digits or as a series of distances to be moved, but because the intended outcome is identical regardless of which encoding of the messages is used, why should the entropy differ?

It is possible to imagine a situation wherein the information necessary to prescribe a movement is greater than the information encoded within the movement (for example, if entropy was lost due to corruption with noise), but this is not what the thought experiment reveals. Figure 5 shows that the information encoded within the fingertip movements is greater than the quantity of information thought necessary to prescribe those motions as described by the index of difficulty – which is troublesome. We are forced to conclude that there is an inaccuracy in the formulation of index of difficulty – it underestimates the quantity of information shown to be encoded within the fingertip movements.

4 The Entropy of a Rapid Aimed Movement

Having concluded that the index of difficulty as defined by Equation 1 is not an accurate measure of entropy, the question that presents itself is, what is the entropy of a rapid aimed movement? The purpose of this section is to address this question using only principles of information theory, with the smallest number of additional assumptions possible. This is not to say that a larger set of assumptions would not have the potential to yield a superior model, but our point here is that a well performing model can be created from an uncomplicated basis, and a small set of assumptions aids in interpreting the model and explaining the existence of Fitts' law.

The model presented here combines three things, an estimation of the entropy pertaining to the distance moved, an adjustment corresponding to the natural variability and inaccuracy of human performance (i.e., errors), and an additional term that accounts for the fact that each movement is a part of an ensemble of movements.

$$ID_{Entropy}\left(D,W\right) = H\left(D\right) - H\left(W\right) + H\left(Ensemble\right) \cdot \tag{15}$$

Note that, so long as the probability distribution of the distances moved (viz., the signal) is statistically independent of the probability distribution of the errors (viz., the noise), then these quantities may be combined through simple subtraction as we have done here (see Shannon [16], page 43, Theorem 16).

4.1 The Entropy Pertaining to the Movement Distance

Entropy provides a measure of the uncertainty within the set of possible outcomes of an information-generating process. So to calculate entropy, we must first identify the set of possible outcomes, and their probabilities. The ultimate goal of a rapid aimed movement is to enter a specific area of space, and so the movement endpoints seem to be the natural quantity of interest that reflects the result of the movement task. In general, the probability distribution of the movement endpoints is unknown.[5] An important detail that arises here is that it is not the software designer's,

[5] Note that the distribution of the scatter of movement endpoints is known to be approximately normally distributed, for example [3, 20]. But this is not the same thing as the distribution of the endpoints, which also depends upon the magnitude of the movement.

experimenter's, nor all knowing being's understanding of the probability distribution that is important, it is the understanding possessed by the user, the experiment participant, or (in the case of the thought experiment) the data entry clerk, that affects their output (see Hyman [4], page 189).

We will consider two possibilities of what someone performing a pointing task may know regarding the probability distribution of movement distances. Either they know the maximum extent of motion, $x \in [0, U]$, (where U is a specific positive distance denoting the size of the movement 'universe'), or nothing $x \in [0, \infty)$. We may estimate the probability distributions corresponding to these assumptions using *principle of maximum entropy*, which states that the probability distribution that best represents a given phenomenon is the distribution yielding the largest entropy, subject to what is known about the phenomenon. The respective entropies are presented in Table 1 (derivations are provided by Shannon [16], pages 35-37). Note that for the case where the maximum distance is not known in advance, $x \in [0, \infty)$, then we must include an additional assumption, which is that the expected distance (i.e., the first moment or mean, of the distribution) is equal to the required movement distance U, which seems reasonable as we would expect the average movement to be centred upon the target, with equal likelihoods of exceeding or undershooting that distance.

Table 1. Maximum entropy distributions for the two cases, where the distribution of maximum movement distances (U) is known by the mover, or not. (Note that conversion from *nats* (natural information units) to *bits* was accomplished by division by $\log_e 2$, see [16, pages 1-2].)

A Priori knowledge	$x \in [0, U]$	$x \in [0, \infty)$
Maximum Entropy Distribution	Uniform Distribution	Exponential Distribution
Entropy (Bits)	$H_{Uniform}(U) = \log_2(U)$ **(16)**	$H_{Exp}(U) = \frac{1}{\log_e 2} + \log_2(U)$ **(17)**

4.2 Noise, Errors, and the Distribution of Movement Endpoints

The distribution of the scatter of movement endpoints is known to be approximately normal [3, 20]. But there is another argument for modelling the scatter of movement endpoints with a normal distribution that can be used without having to explicitly assume that the shape of the distribution is normal. If we presume that errors take the form of deviations in both directions (closer and farther) from the ideal movement distance, then the principle of maximum entropy would have us use the distribution with the maximum entropy defined on the range (-∞, ∞), which is the normal distribution. The entropy of the normal distribution is, (from [16], page 37)

$$H_{Gaussian}(\sigma) = \frac{1}{2} \cdot \log_2\left(2\pi e \cdot \sigma^2\right) \cdot \qquad (18)$$

The relationship between the target width, W, and the standard deviation of the movement endpoints, σ, will affect the error rate. If we assume that the target width represents ±2 standard deviations as Crossman ([2], pages 75-76) suggests it does, then we will match the 4% error rate standard commonly in use today (see [19], page 764). Inserting, $\sigma = {}^{W}/_{4}$, into Equation 18 yields,

$$H_{Errors}(W) = \frac{1}{2} \cdot \log_2\left(\pi e \cdot \frac{W^2}{8}\right). \tag{19}$$

4.3 The Entropy of the Ensemble of Movements

The entropy of an individual outcome depends upon the statistic properties of the set of possible outcomes. For example, consider the entropy of the letter 'A'? If this letter is a part of an ensemble of Unicode UTF-8 characters, then 8 bits of information are needed to represent it. The 'A' could have been any of the other 256 characters of that ensemble, and (see Equation 12), $\log_2 256 = 8$. Alternatively, if the 'A' is part of an ensemble of alphabetic text (26 letters + space) then $\log_2 27 = 4.8$ bits are needed to represent it. Lastly, if the 'A' is part of an ensemble of letters in English text, then it represents between 1 and 2 bits of entropy (see [17]).

Similarly, the entropy of a rapid aimed movement depends upon the statistical properties of the other movements that one may be required to perform. The theoretical implications of this idea, and the practical aspects of accommodating this concept, both deserve more space than we have room for in this paper. Here, we will focus on one particular facet of pointing – Fitts' law does not accommodate the direction of movement, and yet the thought experiment data entry task does. For example, for the $n = 10$ case (where the keyboard contains 10 keys, 0 through 9), if one's finger has just finished hitting the '5' key, and we know the distance to the next key is ±2, there is no way to know which key should be pressed next, '3', or '7'. Ambiguity arises because there are two equally likely movements in the ensemble that must be chosen between. We will work around this issue by adding 1 bit of entropy to account for the directional information (*left* or *right*) that would be otherwise lacking from our model.

4.4 Putting It All Together - The Entropy of a Rapid Aimed Movement

Our model is found by inserting Equations 16 or 17, 19 and the constant 1 into Equation 15. Because the difference between Equations 16 and 17 is a single constant we combine these two cases by introducing a parameter, m,

$$ID_{Entropy}(U, W) = m + \log_2(U) - \frac{1}{2} \cdot \log_2\left(\pi e \cdot \frac{W^2}{8}\right) + 1. \tag{20}$$

where m has a value of 0 if the mover knows the distribution of distances to be moved, or $1/\log_e 2$, (which equals 1.443), if the distribution of distances is unknown.

5 Evaluating the Entropy Formulation of the Index of Difficulty

Reanalysing the thought experiment paradigm using the new entropy-based index of difficulty $ID_{Entropy}$,[6] reveals that over the range of difficulty values examined (1-5 bits), and particularly for larger values (of n and ID), it fits the real (Shannon) entropy better than the traditional index of difficulty. The two variations of the movement distance entropy calculation, sandwich the Shannon entropy between two bounds, with the exponential distance entropy above, and the uniform below, see Figure 7. This suggests that the assumptions and approach are on the right track, but future work remains to determine the role of the m parameter in Equation 20 (is it valid to allow m to take values between the two extremes identified?), and to explore the possibility of using other (non-maximum entropy) distributions.

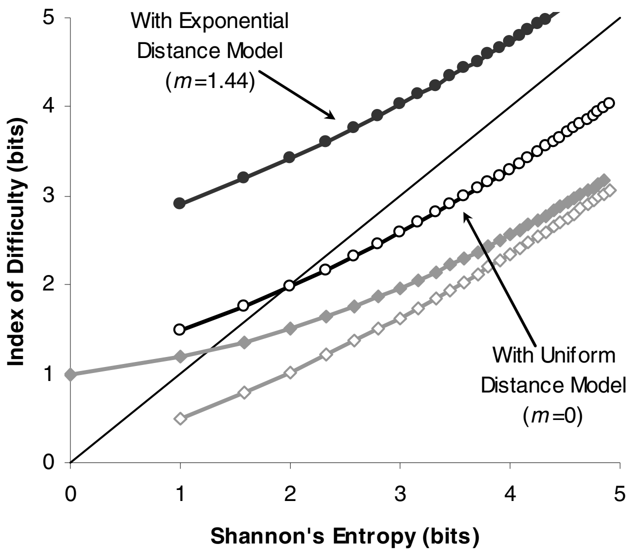


Fig. 6. This figure shows the results of applying the new index of difficulty formulation to the thought experiment results for keyboards with 2 to 30 keys. Two variations of the new model are depicted in black as indicated in the figure. The ordinate of the curve is the new model defined by Equation 20, the abscissa is the entropy found by Equation 12. The line through the origin denotes the position where the model should lay if there was equality between them. The two conditions evaluated previously, *with key repeats* and *without*, appear as the grey curves with empty diamond data points and filled diamonds, respectively.

Figure 7 presents the empirical data from Fitts' 1954 paper [3], analysed with the new and traditional formulations for index of difficulty.

---

[6] The concept, in Equation 20, of the size of the movement universe, is different from the usual Fitts' law distance. It was taken to be the distance between the farthest extents of the *to* and *from* keys of the keyboard (i.e., U=D+1) representing the largest valid key-to-key movement.

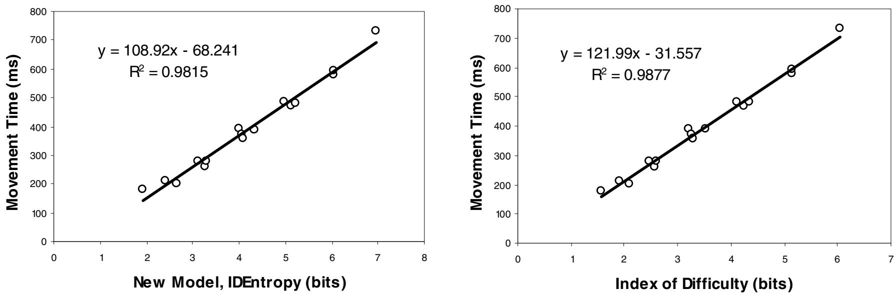


Fig. 7. These figures present an analysis of the empirical data provided by Fitts in his 1954 paper [3], for the 1oz stylus movement task, analysed two ways. On the left we present an analysis of Fitts' data using the entropy formulation of the index of difficulty, Equation 20, assuming the uniform distance entropy variant ($m = 0$). (Note that the value of the m parameter does not affect the fit of the model to the data, it only changes the intercept.) On the right we present an analysis using the standard formulation of the index of difficulty, Equation 1. In both cases the adjustment for accuracy was applied. Note that the R^2 values are similar.

6 Conclusions

The primary contribution made by this paper is the thought experiment that demonstrates the inconsistency between entropy and the Shannon formulation of the index of difficulty. The difference between the index of difficulty and entropy (as depicted in Figure 5) is systemic but not catastrophic, which likely explains why this disparity has not been noticed before, and why many researchers have been able to successfully use Fitts' law to comparatively evaluate the performance of pointing devices in the past. And no doubt the Shannon formulation of the index of difficulty will continue to be used in this fashion in the future.

The problem that has been illuminated here will primarily be of concern when the absolute quantification of information is desired, either to compare pointing devices with other devices for which the true entropy of interaction is known, or to measure human performance in absolute terms (i.e., "How many bits/s can a human output?").

This work also supports the continued use of the adjustment for accuracy in Fitts' law analyses, because performance cannot be measured in terms of entropy without accounting for the equivocation (the bits lost due to noise or errors). For the index of difficulty to be analogous to entropy, errors will have to be similarly accounted for.

Additionally, this paper makes progress toward developing a new entropy-based index of difficulty, that (1) is theoretically rigorous, (2) attempts to resolve the problem demonstrated by the thought experiment, and (3) provides the customary good fit to empirical data, as well as the traditional formulation of Fitts' law does. A small set of assumptions rooted in information theory were used in constructing the entropy *ID* (which with some mathematical rearrangement resembles the traditional formulation), thus implying that what gives rise to Fitts' law is indeed a limitation of human information processing, as opposed to mechanical or biological factors.

Refining the model, exploring the effect that the ensemble has upon the entropy of an individual movement task, and deciphering the parameter m, remains future work.

Acknowledgements. The authors are grateful to the reviewers for their thorough, copious, and very helpful criticism and comments during the review process. This study has been partially supported by Grant-in-Aid for Scientific Research (No. 23300048) in Japan.

References

[1] Card, S.K., Moran, T.P., Newell, A.: The psychology of human-computer interaction. Lawrence Erlbaum, Hillsdale (1983)
[2] Crossman, E.R.F.W.: The Measurement of Perceptual Load in Manual Operations, PhD Thesis, University of Birmingham, UK (1956)
[3] Fitts, P.M.: The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology 47(6), 381–391 (1954)
[4] Hyman, R.: Stimulus information as a determinant of reaction time. Journal of Experimental Psychology 45, 188–196 (1953)
[5] ISO, Ergonomic requirements for office work with visual display terminals (VDTs) – Part 9: Requirements for non-keyboard input devices (ISO 9241-9). Reference Number: ISO 9241-9:2000(E). International Organization for Standardization (2002)
[6] Kvålseth, T.O.: Note on information capacity of discrete motor responses. Perceptual and Motor Skills 49, 291–296 (1979)
[7] Lai, S.C., Mayer-Kress, G., Sosnoff, J.J., Newell, K.M.: Information entropy analysis of discrete aiming movements. Acta Psychologica 119, 283–304 (2005)
[8] MacKenzie, I.S.: A note on the information-theoretic basis for Fitts' law. Journal of Motor Behavior 21, 323–330 (1989)
[9] MacKenzie, I.S.: Fitts' law as a research and design tool in human-computer interaction. Human-Computer Interaction 7, 91–139 (1992)
[10] MacKenzie, I.S., Soukoreff, R.W.: Text entry for mobile computing: Models and methods, theory and practice. Human-Computer Interaction 17, 147–198 (2002)
[11] Meyer, D.E., Smith, J.E.K., Kornblum, S., Abrams, R.A., Wright, C.E.: Speed-accuracy tradeoffs in aimed movements: Toward a theory of rapid voluntary action. In: Jeannerod, M. (ed.) Attention and Performance XIII, pp. 173–226. Lawrence Erlbaum, Hillsdale (1990)
[12] Plamondon, R., Alimi, A.M.: Speed/accuracy trade-offs in target-directed movements. Behavioural and Brain Sciences 20, 279–349 (1997)
[13] Schmidt, R.A., Lee, T.D.: Motor Control & Learning: A behavioral Emphasis, 4th edn. Human Kinetics, Champaign (2005)
[14] Seow, S.C.: Information theoretic models of HCI: A comparison of the Hick-Hyman law and Fitts' law. Human-Computer Interaction 20, 315–352 (2005)
[15] Shannon, C.E.: A mathematical theory of communication. The Bell System Technical Journal 27, 379–423, 623-656 (1948); Republished and easier to obtains as [16], below
[16] Shannon, C.E.: A mathematical theory of communication (1998), Freely available online from: `http://cm.belllabs.com/cm/ms/what/shannonday/paper.html`, Bell Labs
[17] Shannon, C.E.: Prediction and entropy of printed English. Bell System Technical J. 30, 50–64 (1951)
[18] Soukoreff, R.W., MacKenzie, I.S.: Using Fitts' Law to Model Key Repeat Time in Text Entry Models. Poster presented at Graphics Interface – GI 2002 (2002)
[19] Soukoreff, R.W., MacKenzie, I.S.: 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 (2004)
[20] Welford, A.T.: Fundamentals of Skill. Methuen, London (1968)

The Difference Matters: Benchmarking Visual Performance of a Cognitive Pilot Model

Florian Frische and Andreas Lüdtke

OFFIS Institute for Computer Science, Escherweg 2,
26121 Oldenburg, Germany
{frische,luedtke}@offis.de

Abstract. In this paper we introduce an approach to objectively validate visual performance of a cognitive pilot model using benchmarks of human performance. A study with 16 human airline pilots and two competing models has been conducted in order to validate visual performance of the models applying these benchmarks. The study shows that human performance benchmarks can support analysts with a powerful and easy to use method for validation of human performance models. The benchmark is part of a larger-scale method, which will be developed in order to evaluate human factors issues of future HCI-concepts in early stages of system design.

Keywords. Human Performance Modelling, -Validation, -Analysis, Goodness-of-Fit Measures.

1 Introduction

The European project HUMAN[1] aims at developing virtual pilots, in order to improve the human error analysis of new cockpit systems. The virtual pilots should allow simulation-based testing of new cockpit system designs in early design phases. This enables simulating a huge number of scenarios in accelerated time in order to identify potential problems in the operation of new systems (e.g. human errors due to *clumsy automation* [9]) and to derive necessary improvements. The virtual pilots in HUMAN are based on a cognitive architecture named CASCaS (Cognitive Architecture for Safety Critical Task Simulation) [6] which is an implementation of the information processing paradigm of human cognition (similar to ACT-R [1] and MIDAS [5]).

Pilots flying modern aircrafts are confronted with many information (e.g. weather conditions, flight plans, air traffic and automation modes) that are mainly presented on visual displays. Consequently, pilots' vision is the main cognitive resource (apart from auditive and haptic resources) for gathering information about the aircraft and the outside world. Thus, a sophisticated and thoroughly validated visual perception model is a pre-condition for simulating pilot-like interaction with visual interfaces.

---

[1] 7th Framework Programme, see http://www.human.aero for further information about the project.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 240–247, 2011.

Evaluating the predictive power of visual performance models is very complex and time consuming because human visual performance (1) is very variable between different pilots as well as within a single pilot and (2) is a combination of different aspects, such as glance duration, glance frequency and scanpaths. Methods, techniques and tools are needed that can easily be applied by analysts to evaluate model performance.

In HUMAN we validated the performance of CASCaS by comparing data produced by the model with data produced by human pilots in experimental simulator studies, both flying the same scenarios. We developed a benchmark approach that is used in combination with traditional validation techniques. Benchmarks are commonly understood as an analysis method to objectively compare characteristics of subjects with characteristics of a reference subject. The measures considered for the analysis usually characterize the overall power of the system with regard to a target question [10]. Our benchmark approach provides a catalogue of objective measurement criteria allowing (1) to compare the fitness of competing models, (2) to find the best fitting model and (3) to decide objectively if the predictive power of a model is sufficient for the desired application.

In this paper we will briefly introduce our benchmark approach, which we call *Human Performance Benchmarking* and present results of two CASCaS versions that have been compared in order to find the best fitting model. The paper is organized as follows: In Section 2 we present related work in the area of human model validation. Next, in Section 3 our approach will be introduced. In Section 4 we present exemplary results of a first application of the Human Performance Benchmarking. Finally, in Section 5, we will summarize the paper and point out potential improvements of our approach, identified based on the study results. The Human Performance Benchmarks are part of a larger-scale method, on which we are currently working in order to evaluate human factors issues of future HCI-concepts in early stages of system design.

2 Related Work

Different measures, mainly of descriptive statistics, have been used in order to demonstrate the fit of model performance to human performance based on quantitative data (see e.g. [3], [7] and [2]). These measures can be categorized in three types:

1. Measures of central tendency
2. Measures of dispersion
3. Measures of association

The most common type of measure used in the area of model validation are measures of central tendency, e.g. mean, median and mode. Measures of central tendency describe the center or middle of a given data distribution by mapping multiple performance values on a single value. Thus, measures of central tendency are not sufficient to describe the variability in human performance.

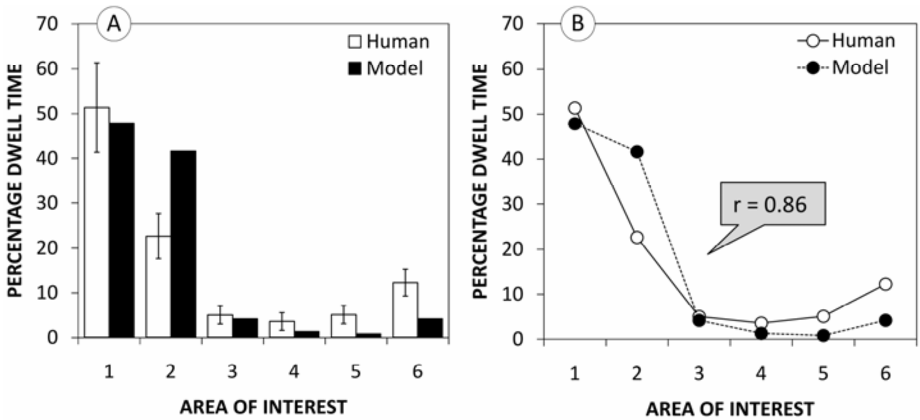


Fig. 1. Comparison of glance distribution of pilots in a flight simulator study conducted at German Aerospace Center (DLR). Chart A depitcts a common bar chart used to visualize local fitness including the mean and standard deviation for each area of interest. Chart B depicts the mean glance distribution as line chart (which is more intuitive for analysing trend consistency) including Pearson (r) as a quantification of trend fitness.

However, variability of pilot behavior has to be considered during cockpit design, because the design has to be safe and usable for different types of pilots. Measures of dispersion are used to describe the between-subjects variability and the within-subject variability. Typical measures of dispersion are range, standard deviation and confidence intervals. Finally, measures of association are used (1) to describe the relation of data points of a sample to data points of another sample, or (2) to describe the relation between parameters within a sample. Typical measures of association are Pearson's correlation coefficient (in the following referred to as *Pearson*) and Spearman's or Kendall's rank order correlation. All three types of measures are frequently used to validate models of human performance with different strength and weaknesses.

Measures of central tendency in combination with measures of dispersion are used to describe the local fit between human and model performance data, see e.g. [7] and [3]. Combining mean (as measure of central tendency) and standard deviation (as measure of dispersion) allows to quantify the average case and variability of human and model behavior based on given data samples. Fig. 1A shows an example for measuring distribution of gaze on areas of interest. The combination of mean and standard deviation provides a *qualification* of local fitness between datasets. Nevertheless, they do not *quantifiy* fitness. Quantification is especially needed if analysts have to validate different versions of a model in order to find the best fitting one.

Measures of association are used to quantify the relation between model data and human data by calculating the trend fitness between the two datasets. Fig. 1B shows an example for comparing the rank order of the pilots' gaze distribution. Here, Pearson is used to quantify the consistency of the rank order for the two data sets. Nevertheless, measures of association do not take into account the variability within human performance. Quantifications calculated based on measures of association provide only poor evaluation of the fitness between model and human data because they do not consider behavioural variability. It is not possible to draw any conclusion about the final predictive power of the model. Thus, a measure is needed which on the one hand considers variability in human performance and on the other hand quantifies model performance in comparison to human performance.

First approaches to provide such a measure have been researched by Schunn and Wallach, and by Gluck. Schunn and Wallach provided a measure for quantification of local fitness in [8]. They recommend using the root mean square deviation (RMSD) as a measure of local fitness. RMSD calculates the difference between a set of related performance values and merges them on a single value of predictive power. Thus, RMSD allows comparing performance of different variants of a model on a quantitative basis. However, the problem analysts are faced with when using this approach for validation of local fitness is analog to the problem when using Pearson, that is the absence of taking into account human performance variability.

This problem has been tackled by Gluck et al. in [4]. They proposed an approach to validate a pilot model, taking into account the between-subject *differences* of a set of human performance datasets instead of absolute performance values. The between-subject differences served as a reference in order to actually evaluate the goodness-of-fit of a pilot model *m*. For a group of human subjects *S*, they calculated the difference in performance of each individual subject s_k to the average performance of $S \backslash s_k$ using RMSD. The result was a set of difference values describing the natural variability of human performance which has been used to evaluate the difference between *m* and *S*. When analysing visual data, one of the characteristics most frequently described is the glance distribution on a set of pre-defined areas of interest. Here, RMSD for local fitness and Pearson for trend fitness can be used to calculate the difference between two subjects. Consequently, for a group of *n* human subjects, the result can be described as a bi-directional complete graph K_n (see Fig. 2) where *K* denotes the graph and *n* depicts the number of nodes. Edges $i \leftrightarrow j$ are labeled with a value v_{ij} representing the difference (RMSD or Pearson) between s_i and s_j. Thus, the graph completely describes the differences between human subjects (difference graph). The same approach can be used for measuring differences between human subjects and human performance models.

Human Performance Benchmarks use difference graphs as the basis in order to validate visual performance of human models tackling quantification of visual performance and taking behavioural variability into account.

3 Human Performance Benchmarking

Benchmarks are commonly understood as an analysis method objectively comparing characteristics of subjects with characteristics of a reference subject. The measures considered for the analysis usually characterize the overall power of the system with regard to a target question [10]. A benchmark is a quantitative test evaluating the characteristics observed and matching results on a scale which is independent from the characteristics. Thus, results of different characteristics can be aggregated and evaluated on different levels of abstraction. This allows analysts to identify strengths and weaknesses of systems objectively based on data.

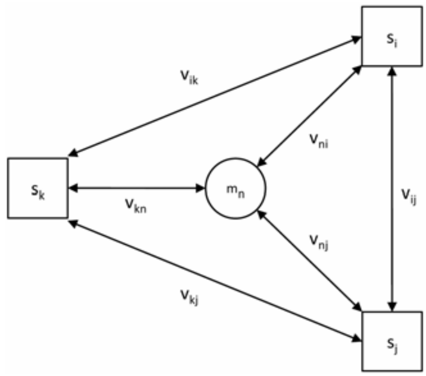


Fig. 2. Difference graph describing the individual performance differences of human subjects (square) to each other as well as the differences between human subjects and a model (circle), or between competing model versions

The intention of our benchmark for visual performance is providing a goodness-of-fit evaluation for different characteristics C (such as glance distribution, glance frequency or scanpaths) for a set of competing models M with regard to a set of human subjects S. Our benchmark algorithm calculates an individual fitness value for each model $m \in M$. Fitness values can be used (1) to compare the fitness of competing models, (2) to find the best fitting model and (3) to decide objectively if the predictive power of a certain model is sufficient for the desired application. In the following, the algorithm used to calculate a model's fitness value is presented:

1. For all $c \in C$, calculate difference graph for m and S
2. For all $c \in C$, calculate confidence interval CI_S of individual difference values in S
3. For all $c \in C$, calculate the mean $mean_m$ of differences between m and S
4. For all $c \in C$, evaluate similarity between m and S
5. Calculate overall fitness value fit_m of m based on similarity evaluations of each c

The similarity evaluation in step 4 can be defined as a function $sim_c(m,S)$ evaluating for each c, if the performance of a given m is human-like, where human-like performance is determined by S. Our first version of this function is a naive boolean function returning *true* if $mean_m$ is covered by CI_S and *false* if $mean_m$ is not covered by CI_S:

$$sim_c(m,s) = \begin{cases} true \mid mean_m \in CI_S \\ false \mid else \end{cases} \tag{1}$$

In step 5, fit_m is calculated. We define fit_m as the sum of *true* similarity evaluations divided by the sum of all (*true* or *false*) similarity evaluations. Thus, the range of fit_m

is [0;1], where 0 means no fitness and 1 means total fitness. Based on fit_m, analysts can decide if a model's performance is sufficient for the desired application or not. A model is sufficient, if $fit_m \geq thres$, where *thres* is a pre-defined threshold parameter.

The application of this algorithm allows considering performance variability in terms of between-subject differences provided by difference graphs (step 1) which are used to calculate an interval CI_S which represents the variability (step 2). In addition, the algorithm allows quantification of fitness by applying $sim_c(m,S)$ (step 4) and calculating fit_m (step 5) as a measure of model fitness. In the following Section, we will present exemplary validation results of an application of Human Performance Benchmarks in order to preliminary assess the approach.

4 Results

Experiments with 16 human airline pilots have been conducted in order to collect reference data for the validation of the visual performance data of our pilot model.

Performance data of human pilots and two competing model versions (m_1 and m_2) have been validated based on data of glance distributions during three flight phases (cruise (1), approach (2) and final approach (3)), see Fig. 3A.

Accordingly, the benchmark has been used to quantify performance and to objectively evaluate fitness of m_1 and m_2 for three performance characteristics c_1 (glance distribution during cruise phase), c_2 (glance distribution during approach phase) and c_3 (glance distribution during final approach phase):

Step 1: We computed the corresponding difference graph for c_1, c_2 and c_3 taking into account the 16 subject pilots S and m_1 and m_2. Due to size limitations the graphs are not shown here.

Step 2: We computed CI_S for c_1, c_2 and c_3. The results are shown in Fig. 3B.

Step 3: We calculated $mean_{m1}$ and $mean_{m2}$. The results are shown in Fig. 3B.

Step 4: We evaluated the similarity between m_1 and S, and between m_2 and S. The results are shown in Table 1. According to the similarity function defined in Section 3, Fig. 3B shows that only m_1 receives a *true* evaluation ($sim_{c2}(m_1,S) = true$).

Step 5: Results of step 4 have been used to calculate the overall fitness value of m_1 and m_2. The results are shown in Table 1. For the datasets analysed in this study, the overall fitness value of model 1 is .33 and the evaluation of model 2 is 0. The fitness values can be used to evaluate if these models are sufficient for the desired application based on a comparison of each fitness value to *thres*. (e.g., if analysts define $thres = .3$, then m_1 would be sufficient).

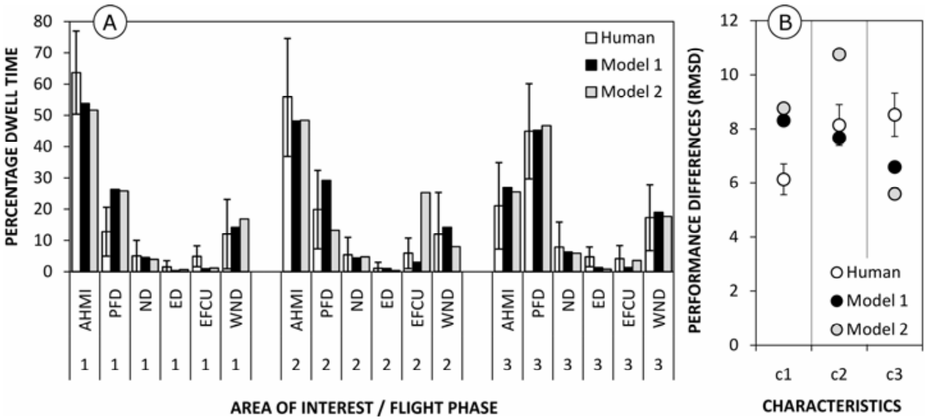


Fig. 3. Chart A depicts the glance distributions (AOIs: Advanced Human Machine Interface (AHMI), Primary Flight Display (PFD), Navigation Display (ND), Engine Display (ED), Electronic Flight Control Unit (EFCU) and Window (WND)) of a human population (mean and standard deviation) and two concurring models for three flight phases (cruise (C), approach (A) and final approach (F). Chart B depicts the mean values and confidence intervals (95\%) of human pilots and the mean values of model 1 and model 2, which have been calculated based on the difference graphs for the flight phases.

Table 1. Results of benchmarking reveal that m_1 receives a higher overall fitness value than m_2

Model	Cruise	Approach	Final Approach	Overall Fitness Value
m_1	false	true	false	.33
m_2	false	false	false	.0

5 Summary and Future Work

In this paper, we presented an approach (called *Human Performance Benchmarking*) for quantitative evaluation of goodness-of-fit for human performance models based on quantitative performance data. The approach considers variability in human performance and uses these performance differences as a model evaluation reference. A first naive evaluation function has been presented, which evaluates the model fitness in context of these reference data. The benchmarking approach has been applied within a study with aircraft pilots focusing on gaze data. The results have been presented and issues requiring improvements have been identified. The first issue is that the evaluation function is very simple and results may be misleading for models that perform just within the human performance intervals. Evaluation of these models would be similar to those that perform much better and would considerably differ from models that perform slightly worse. We plan to improve the evaluation function by weighting the distance to the center of the distribution. The second issue is that the function works fine for deterministic models (same input - same output) but models like CASCaS produce performance variance as well. We have not yet taken variability in model performance into account. We plan solving this issue by considering CI_m instead of $mean_m$ for a set of model runs and to compare CI_m to CI_S.

Acknowledgments. The work described in this paper is funded by the European Commission in the 7th Framework Programme, Transportation under the number FP7 - 211988.

References

1. Anderson, J.R., Bothell, D., Byrne, M.D., Douglass, S., Lebiere, C., Qin, Y.: An integrated theory of the mind. Psychological Review 111, 1036–1060 (2004)
2. Ball, J.T., Gluck, K.A., Krusmark, M.A., Rodgers, S.M.: Comparing three variants of a computational process model of basic aircraft maneuvering. Institute for Simulation and Training, 87–98 (2003)
3. Frische, F., Osterloh, J.P., Luedtke, A.: Modelling and validating pilot visual attention allocation during the interaction with an advanced ight management system. In: Human Modelling in Assisted Transportation: Models, Tools and Risk Methods, pp. 165–172. Springer, Heidelberg (2011)
4. Gluck, K.A., Ball, J.T., Krusmark, M.A., Rodgers, S.M., Purtee, M.D.: A computational process model of basic aircraft maneuvering. In: Proceedings of International Conference on Cognitive Modelling (ICCM), pp. 117–122. Universitaets-Verlag, Bamberg (2003)
5. Gore, B.F., Corker, K.M.: Increasing aviation safety using human performance modeling tools: An air man-machine design and analysis system application. In: Chinni, M.J. (ed.), pp. 183–188 (2002)
6. Lüdtke, A., Osterloh, J.-P., Mioch, T., Rister, F., Looije, R.: Cognitive modelling of pilot errors and error recovery in flight management tasks. In: Palanque, P., Vanderdonckt, J., Winckler, M. (eds.) HESSD 2009. LNCS, vol. 5962, pp. 54–67. Springer, Heidelberg (2010)
7. Salvucci, D.D.: Modeling driver behavior in a cognitive architecture. Human Factors 48 (2006)
8. Schunn, C.D., Wallach, D.: Evaluating goodness-of-fit in comparison of models to data (2005)
9. Wiener, E.L.: Human factors of advanced technology (glass cockpit) transport aircraft. Tech. rep., NASA Ames Research Center (1989)
10. Zhang, X.: Application-specific benchmarking. Tech. rep., Harvard University (2001)

Visual Search in Radial Menus

Krystian Samp and Stefan Decker

Digital Enterprise Research Institute,
National University of Ireland, Galway
{krystian.samp,stefan.decker}@deri.org

Abstract. Menu research has focused predominantly on linear menus (e.g., cascading menus). Little is known about user behavior with radial menus, which have been around for some time. The paper investigates the order in which users find items in radial menus. We analyze data collected in a controlled experiment and define serial position for items laid out in a circular fashion. For the first level (ring), the serial positions start at 12 o'clock position and alternate between both sides of the ring. For subsequent levels, the serial positions follow distance from a parent item. The defined search pattern yields strong fit and has substantial effect on search performance. We discuss the results in the context of radial menu design.

Keywords: Visual search, search patterns, radial menus, pie menus.

1 Introduction

Menus are a primary GUI technique and subject of considerable amount of research. One of the important areas of that research is how novice users search the menus. Although search strategies vary across the users [1], eye tracking studies demonstrate that search is predominantly top-to-bottom [4]. Consequently, top menu items are found faster (experiment 2 in [12]). This finding contributes to the understanding of user behavior with menus, the development of cognitive and menu performance models [5, 7], but also to improvements in menu design. For example, menu access can be improved by arranging items so the most important, frequently selected [14], recently selected [8], or user-specified [8] items are placed in the top of the menu for optimal search and selection.

Previous research focused predominantly on linear menus and similar findings regarding menu search are not available for radial menus. The findings regarding visual search in linear menus cannot be readily generalized to radial menus as the layout differences are considerable. Consequently, the question of how to arrange items in radial menus for optimal access remains unanswered. This question gains weight given the popularity of radial menus in research [2, 6, 11, 13], professional applications (e.g., Autodesk SketchBook Pro 2010), and computer games (e.g., Rockstar Red Dead Redemption 2010). The objective of this paper is to make the first step toward understanding how users search radial menus.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 248–255, 2011.

2 Method

There are a number of plausible strategies for how the items on the rings could be searched – for example, clockwise, counter-clockwise, top-to-bottom or bottom-to-top. The search could start from an item closest to the parent item or from an item on any end of an arc (see Figure 1, FAN and CRL-SMALL designs). Search might also lack any systematic patterns and be best described as random. Note, that the different search strategies result in different items being found first. Therefore, it is not clear how to arrange items for optimal menu access and it is necessary to analyze empirical data.

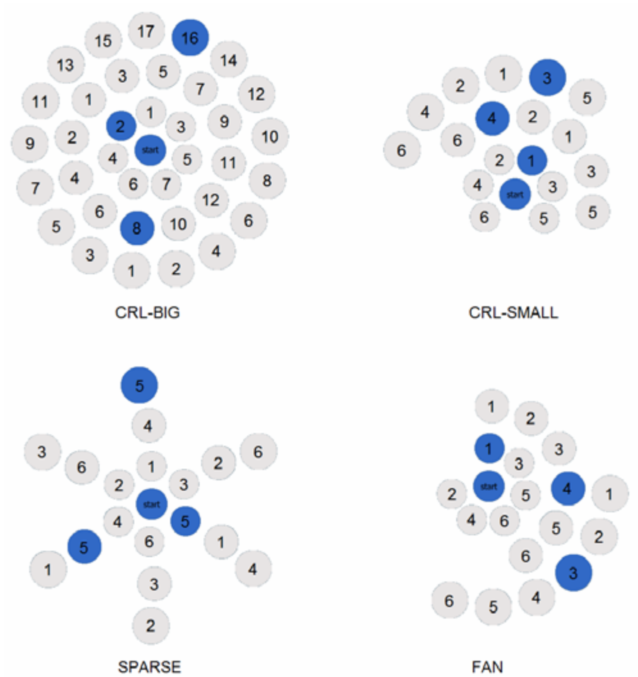


Fig. 1. Radial menus used in the experiment presented in [13]. The labels have been removed and substituted with serial positions. The selected items are colored blue.

We use the data collected in the experiment presented in [13]. In that experiment, 18 participants performed selections from one linear and four radial menus. The radial menus are presented in Figure 1.

The menus use concentric rings, one for each level, to organize the items. The most inner ring corresponds to the first level. The SPARSE design distributes items evenly within each ring. The FAN design places items clockwise around the ring starting from the top (for the first ring) or from the position closest to the parent item (for all subsequent rings). The Compact Radial Layout (CRL) design packs items tightly around their parent, keeping the distances between items and levels as short as possible. The behavior of the radial menus is similar to cascading menus. A user is initially presented only with the first level. A click on menu items post and un-post further levels (rings).

The menus used in the experiment had three levels and were of two different sizes: 1) small having 6 items on each level (CRL-SMALL, SPARSE and FAN in Figure 1); 2) big having 7, 12 and 17 items fully populating first, second and third level respectively (CRL-BIG in Figure 1). Menu content was randomly generated for each trial to prevent any learning effects and enforce visual search (i.e., to simulate novice behavior). A dictionary of nouns and adjectives was used to make the selection sequences easy to remember.

Visual search time was separated from pointing time using the Point-Completion Deadline (PCD) technique described in [10]. In short, the technique discourages participants from moving the mouse until they found the target. Consequently, the start of the mouse movement consistently marks the end of visual search. The PCD is an accepted, legit technique used to reliably separate and measure visual search times in selection tasks [3, 10, 13]. This technique, however, does not employ eye-tracking. Therefore, what we focus on in this paper is not patterns of the actual eye movements but rather patterns identifying **order of finding** the items. Such patterns would inform design of radial menus as to how order the items for optimal search, would shed some light at how users search radial menus, and would help formulate hypotheses for more detailed eye-tracking studies.

The decision to analyze the data from [13] was motivated by a number of reasons: 1) the designs presented in [13] resemble many radial menus used in production, such as those mentioned in the introduction; 2) the experiment in [13] included three different radial menus, two menu sizes, and three levels, enabling broader validation of our investigation; 3) the first levels of the designs in [13] are similar to the Pie Menu [6] in terms of searching; 4) the SPARSE design is similar to the Wave Menu 2 [3]. We hope these reasons will assure broader applicability of the results.

The authors in [13] reported average times for each menu design and each menu level. Therefore, from their study it is not known how position on the ring affects measured times or which items are found faster.

To find regularities in how participants search the rings, we aggregated visual search time across participants using an arbitrarily chosen but consistent scheme of numbering the items around the rings. This was done for each menu design and each menu level separately. Ordering the items according to the visual search times led to an observation that for the second and third levels, search time seemed to be a function of the distance from a parent item. For the FAN design this results in search following the ring clockwise starting from the item closest to the parent. For the CRL-BIG, CRL-SMALL, and SPARSE designs the result is search alternating between both sides of the ring (i.e., half rings on both sides of the parent item), again starting from the item closest to the parent (consult with Figure 1). For the first level, where all the items are equally distant from the center, we observed that search starts from the top item (i.e., the item at 12 o'clock) and also alternates between both sides (in this case left and right) of the ring.

2.1 Results: Analyzing General Search Pattern

Consistent with our observations, we define serial position of the items on the first ring in an alternating manner starting from the top item (i.e., the item at 12 o'clock). For all subsequent rings, serial position of the items is defined according to their distance from a parent item. Figure 1 superimposes serial positions on the items of the four radial menus. From now on, whenever we say the first or the last items, we refer to the order defined by the serial position.

We performed regression analysis to measure the actual relationship between the serial position and the visual search time. For each menu design and serial position, times were aggregated across participants and levels. It allowed us to obtain a number of observations necessary for establishing reliable mean time per serial position. Moreover, it allowed us to focus on the general pattern of which items are found faster regardless of menu level. If such a general pattern exists, it will be more readily applicable in practice compare to, for example, different patterns for different levels. We also performed a separate regression on times aggregated only across small radial menus.

Note that the aggregation across levels could flatten the possible effect of level, especially for the big radial menu, where each level contains a different number of items. To this end, we also analyze the effect of menu size later in the paper.

2.1.1 Visual Search as a Function of Serial Position

The average number of observations per serial position was 133 (SD 30) for small radial menus (CRL-SMALL, SPARSE, FAN) and 97 (SD 60) for big radial menu (CRL-BIG). The results are shown in Figure 2. The fit is strong across menu designs and two menu sizes (R^2>.85 and R^2=.99 for times aggregated across three small radial menus). The results indicate that the defined search pattern holds for all menu designs used in the experiment. Interestingly, search alternating between both sides of the ring for the CRL-SMALL, CRL-BIG, and SPARSE designs (being a function of the distance from a parent item) leads to a non-optimal scanning path in terms of its total length. Search following circumference along one direction (e.g., clockwise starting in any position) would have resulted in a shorter path. We hypothesize that, not being able to find a target item in a given location, participants felt encouraged to move to an entirely different location.

Regression lines are fairly similar across small radial menus (intercepts ranging from 1044 ms to 1077 ms and slopes ranging from 61 ms to 76 ms). Each subsequent serial position is associated with an average of 70 ms increase of visual search time. Regression line for the big radial menu has notably larger intercept (1410 ms) and slope (100 ms) when compared with the results for the small menus. For the big menu, each subsequent serial position is associated with an average of 100 ms increase of visual search time. It indicates that the menu size (length) affects search performance. We explore this in the next section.

Figure 2 shows that the difference between finding the first and the last item is substantial. For small radial menus, this difference is on average 360 ms (approximately 30% of the average visual search time for small menus). For the big radial menu the difference is 1900 ms (approximately 85% of the average visual search time for the big menu). Consequently, arranging items on the rings can lead to notable performance differences.

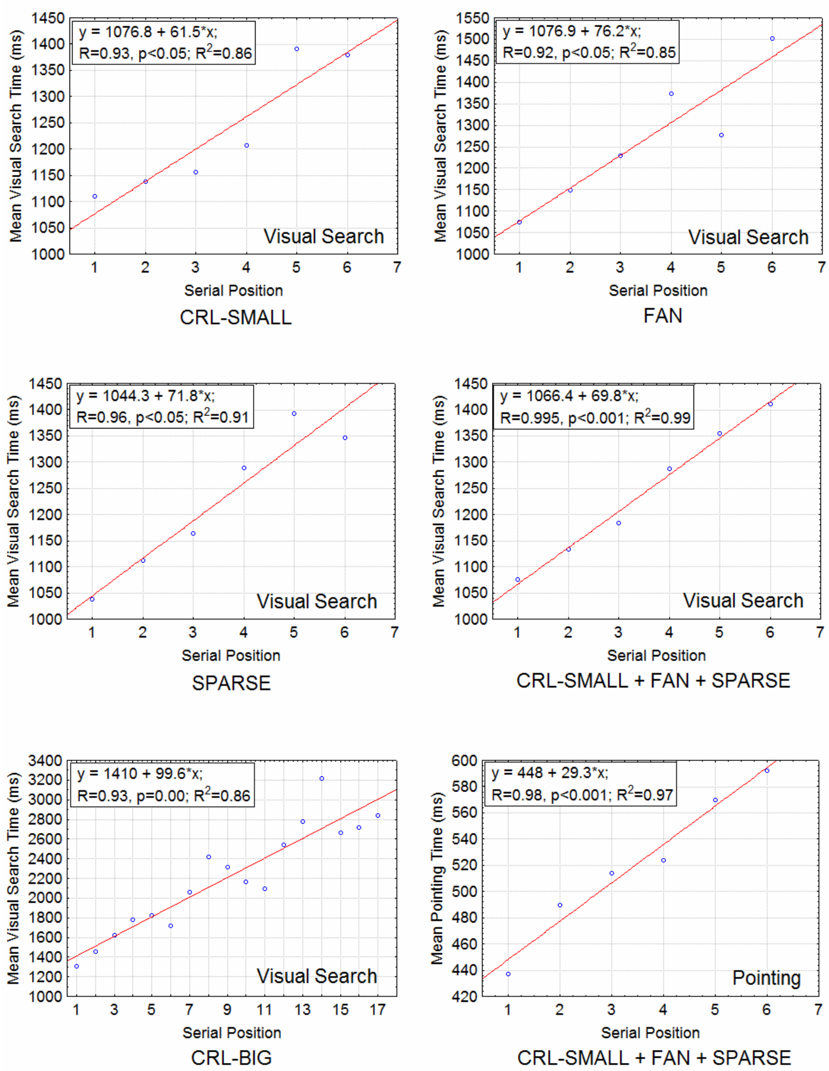


Fig. 2. Visual Search Time and Pointing Time as a function of serial position. The best-fit regression lines and corresponding statistics are provided for each condition. Notable dispersion of the visual search times after the seventh item for the CRL-BIG menu can be partially attributed to the smaller number of observations available for those item positions.

Note also that the defined serial position will also have linear effect on pointing time since each subsequent item (according to serial position) is more distant from a parent item and thus slower to select according to Fitts' law [9]. We run regression analysis of pointing times against serial position (see the last graph in Figure 2). As expected, the fit is strong (R^2=.97 for times aggregated across the three small radial menus, and R^2=.81 for the big radial menu). Restricted by the available space, we did not include a graph for the big radial menu. Its lower correlation may be partially attributed to a smaller number of observations available for that menu. For small radial menus the difference between selecting the first and the last item is 160 ms (approximately 30% of the average pointing time for small menus). For the big radial menu, the difference is 300 ms (approximately 48% of the average pointing time for the big menu). Consequently, arranging items according to the proposed pattern (i.e., most important items in the first serial positions) not only will improve search time but also pointing to these items which is important for both, novices and experts [7].

2.1.2 Visual Search as a Function of Menu Size

The big radial menu (CRL-BIG) has different number of items on each level (7, 12, 17). This allowed us to measure the effect of menu size on visual search time much like [12, 14] did for linear menus. We aggregated times for the seven first items on each level and performed regression between these times and the three menu sizes. Regression analysis produced a strong model with SearchTime = 140*MenuSize + 209, R^2=.99. The visual search time increases by approximately 700 ms between the analyzed menu sizes[1].

Menu size has similar linear effect on visual search time for radial menus as it has for linear menus. However, the size of the effect is notably larger for radial menus. Our results indicate that each additional item in radial menu increases average visual search time by 140 ms. Cockburn et al. [7] report 80 ms increase for each additional item in linear menu. The difference is considerable. It would be interesting to explore which parameters of radial menus – for example, circular trajectory, item density, item spacing, item shapes – have the strongest effect on the slowdown.

2.2 Results: Analyzing Individual Search Patterns

Following our investigation, we analyzed individual search patterns. For each participant, menu design, and serial position, times were aggregated across the levels and plotted as a function of serial position (72 plots). We made two observations analyzing the individual plots. First, sometimes participants searched not one, but two or three consecutive items before moving to the opposite side of the ring. Second, participants employed not one but various search strategies, of which the following were dominant: 1) the search pattern described in the paper; 2) searching in an alternating manner but starting with the last items; 3) starting with the first items and searching half of the ring to the left/right before proceeding to the other half. Some searches could be best described as random. Note that for some of these strategies the items with the first serial positions were actually found the last[2]. For this reason, the pattern reported in our paper should not be treated as a definitive order in which participants find items on the ring but rather as general regularity (much like top-to-bottom regularity in linear menus [1, 4]). In other words, the pattern indicates which items are found faster and slower in a statistical sense – i.e., on average.

---

[1] We also performed the same analysis for four menu sizes (including size=6 from small menus). The results were almost identical, R^2=.98.

[2] We tried alternative ordering schemas but they resulted in a poor fit. For practical reasons and space wise we focused on the prevailing search pattern.

For the same reason, and to better understand strength of the proposed pattern, we decided to extend our analysis and check how often the first items are actually found faster. To this end, for each participant and menu design, we aggregated visual search time for each serial position. Further, for each participant and design, we counted how many items from the first half of all the items (according to our serial position) a participant found faster than the half of all the items (i.e., faster than the majority of the items). For each participant we average this count across the menus to obtain one count per participant – we call this count TOP-IN-TOP. Similarly, we counted how many items from the first half of all the items a participant found slower than the half of all the items. For each participant, the TOP-IN-BOTTOM is the average of this count across the menus. TOP-IN-TOP and TOP-IN-BOTTOM, divided by their sum, represent the probabilities of finding an item from the first half of the items faster/slower than a half of all the items (in other words faster/slower than the majority of the items). The distributions of TOP-IN-TOP and TOP-IN-BOTTOM measures were normal (the Shapiro-Wilk's W test used, $p>0.05$). We performed the dependent measures t-test to determine if TOP-IN-TOP and TOP-IN-BOTTOM differed significantly.

For small radial menus, the mean number of TOP-IN-TOP items was 2 (SD 0.58) (associated probability 67%) and the mean number of TOP-IN-BOTTOM items was 1 (SD 0.58) (probability 33%). For the big radial menu, the mean number of TOP-IN-TOP items was 6.5 (SD 1) (probability 72%) and the mean number of TOP-IN-BOTTOM items was 2.5 (SD 1) (probability 28%).

The results of the t-test show that, indeed, the first half of items determined by our serial position are found faster than the majority of items ($p<.05$ for small radial menus, and $p<.01$ for the big radial menu). This reinforces the results obtained in the regression analysis. However, reported means and probabilities demonstrate that the search pattern determined by the serial position is not as strong as could be asserted solely on the basis of correlation score. In other words, other search patterns are also important

In sum, placing the most important items according to the proposed pattern gives approximately 70% of chances for these items to be found faster than the majority of the items. This is 20% better than a random order. Given substantial differences between visual search time for the items found the fastest and the slowest (30-85%, see section 2.1.1), following the proposed pattern will lead to better overall performance of radial menus. This is important as searching radial menus is known to be slower than searching linear menus [13].

The existence of other patterns hints that designers might consider solutions that suggest the sense of reading, such as using an arrow or using animation which reveals the items in a desired order.

3 Conclusions

In this paper we investigated the order in which the items are found in radial menus. We found that for the second and third levels, search time is a function of the distance from a parent item. For the first ring, where all the items are equally distant from the center, search alternates between both sides of the ring, starting from the top item.

The difference between finding the first and the last item is substantial: 30% and 85% of the average search time in small and big radial menu respectively. Given that the search is a dominant component of menu selection times for novices [7], the order of the items on a ring will affect menu performance.

Our results suggest that the items a designer wishes to be found faster (e.g., most important items) should be placed in the top of the first ring and in the positions closest to the parent item for all subsequent levels. Users will also benefit from decreased pointing times to these items.

The future work could use eye-tracking devices to explore the actual eye movement patterns.

References

1. Aaltonen, A.: Hyrskykari, and A., Räihä, K. J.: 101 spots, or how do users read menus? In: International Conference on Human Factors in Computing Systems (CHI), pp. 132–139 (1998)
2. Bailly, G.: Wave Menus: Improving the novice mode of Marking Menus. INTERACT, 475–488 (2007)
3. Burke, M., Hornof, A.J., Nilsen, E., Gorman, N.: High-cost banner blindness: Ads increase perceived workload, hinder visual search, and are forgotten. ACM ToCHI 12(4) (2005)
4. Byrne, M., Anderson, J., Douglass, S., Matessa, M.: Eye Tracking the Visual Search of Click-Down Menus. In: CHI, pp. 402–409 (1997)
5. Byrne, M.D.: ACT-R/PM and menu selection: applying a cognitive architecture to HCI. International Journal Human-Computer Studies 55, 41–84 (2001)
6. Callahan, J., Hopkins, D., Weiser, M., Shneiderman, B.: An empirical comparison of pie vs. linear menus. In: CHI, pp. 95–100 (1988)
7. Cockburn, A., Gutwin, C., Greenberg, S.: A Predictive Model of Menu Performance. In: CHI, pp. 627–636 (2007)
8. Findlater, L., McGrenere, J.: A comparison of static, adaptive, and adaptable menus. In: CHI, pp. 89–96 (2004)
9. Fitts, P.: The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement. Journal Experimental Psychology 47, 381–391 (1954)
10. Hornof, A.J.: Visual Search and Mouse-Pointing in Labeled versus Unlabeled Two-Dimensional Visual Hierarchies. ACM ToCHI 8(3), 171–197 (2001)
11. Kurtenbach, G.: The Design and Evaluation of Marking Menus. PhD Thesis (1993)
12. Nilsen, E.L.: Perceptual-motor control in human-computer interaction (Tech. Rep. No. 37). Ann Arbor, Michigan: The Cognitive Science and Machine Intelligence Laboratory, The University of Michigan (1991)
13. Samp, K., Decker, S.: Supporting menu design with radial layouts. In: International Conference on Advanced Visual Interfaces (AVI), pp. 155–162 (2010)
14. Sears, A., Shneiderman, B.: Split menus: effectively using selection frequency to organize menus. ACM ToCHI 1(1), 27–51 (1994)

Analytic Trails: Supporting Provenance, Collaboration, and Reuse for Visual Data Analysis by Business Users

Jie Lu, Zhen Wen, Shimei Pan, and Jennifer Lai

IBM T. J. Watson Research Center, 19 Skyline Drive,
Hawthorne, NY, 10532, USA
{jielu,zhenwen,shimei,jlai}@us.ibm.com

Abstract. In this paper, we discuss the use of analytic trails to support the needs of business users when conducting visual data analysis, focusing particularly on the aspects of analytic provenance, asynchronous collaboration, and reuse of analyses. We present a prototype implementation of analytic trail technology as part of Smarter Decisions – a web-based visual analytic tool, with the goal of helping business users derive insights from structured and unstructured data. To understand the value and shortcomings of trails in supporting visual analytic tasks in business environments, we performed a user study with 21 participants. While the majority of participants found trails to be useful for capturing and understanding the provenance of an analysis, they viewed trails as more valuable for personal use rather than for communicating the analytic process to other people as part of a collaboration. Study results also indicate that rich search mechanisms for easily finding relevant trails (or portions of a trail) is critical to the successful adaptation and reuse of existing saved trails.

Keywords: Information visualization, Visual data analysis, Analytic provenance, Asynchronous collaboration, Analysis reuse.

1 Introduction

It is becoming increasingly common for business workers to need to analyze large amounts of data in order to derive the insights necessary for business decisions. Finding an effective way to turn data overload into information that can be used to make decisions quickly has become a high priority [22]. As a result, data analytic tools, particularly those that provide the ability to visualize data with charts, graphs, and maps (i.e. "visual analytic tools"), have attracted increasing attention in recent years [18, 26].

Despite this growing use and acceptance of visual data analysis, several problems exist with current visual analytic tools in business environments. First, the highly interactive and exploratory nature of visual analytic activities often makes it difficult for the user to capture the steps and metadata of the analytic process which are needed to facilitate effective re-visitation [10]. Without adequate support for capturing and retracing the provenance of an analytic process it is difficult and time-consuming to reconstruct or understand how a particular insight was discovered or why a decision was made. Second, with the rise of business globalization, people working together on

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 256–273, 2011.

a task are often separated by time and distance, requiring them to work asynchronously. It is a challenge to use today's tools for effective collaborative visual analysis [14]. Third, the data sets that a business worker needs to analyze at different times often come from the same domain (e.g. sales figures) and require similar types of analysis. Because current tools provide limited support for reusing or adapting pre-existing analyses, the user mostly has to start from scratch each time s/he analyzes a new data set.

In an attempt to address these problems, we have developed the analytic trail technology as part of *Smarter Decisions*, an interactive web-based visual analytic tool built to enable users who are not visualization experts to interact visually with both structured (e.g. relational database, spreadsheets) and unstructured (e.g. paragraphs of text, blogs, articles) data. This technology automatically captures trails of the analytic steps taken by the user during visual data exploration and displays them as an interactive GUI component. Such trails can create a "corporate memory" of the decisions that were made. They can be rolled back at any time to view each step of the analysis, thereby increasing the transparency of decision making (e.g. who made the decision and why). Trails can be shared, allowing teams to collaborate in decision making. Saved trails can also be used as template or model to facilitate new analysis based on existing stored trails. Our goal in developing this technology was to provide support for three key needs of visual data analysis in business environments: analytic provenance, asynchronous collaboration, and analysis reuse.

To evaluate the effectiveness of the analytic trail technology at supporting this goal, we conducted a user study (N=21) of analytic trails as part of the *Smarter Decisions* tool. The study results help to shed light on situations where trails can provide the greatest benefits as well as the design considerations required to achieve these benefits.

In the following sections of the paper, we begin by describing related work. We then present insights gained from semi-structured interviews with six business analysts about their process and requirements when analyzing business data visually, which informed the design of the analytic trail technology. Next, we provide an overview of the *Smarter Decisions* tool, followed by detailed description of the analytic trail technology. Finally, we present results from the user study, and conclude with a discussion of the findings and directions for future work.

2 Related Work

Our work is related to several areas of research including visual analytic provenance, reuse, and asynchronous collaborative visual analytics. In this section, we review key papers in these areas upon which our research builds.

2.1 Analytic Provenance and Reuse

Research has shown that preserving a historical record of visual analytic activities (i.e. provenance) is an important requirement in many visual analytic applications [20, 25]. To capture visual analysis history, researchers have explored the use of various history models, visual representations, and operations. Graph-based [16, 23] and tree-based [1, 2] history models have been developed for capturing complex non-linear analysis history. Taxonomies and classification schemes have been proposed to categorize actions in visual analysis [7, 10, 13, 27, 28]. Depending on the underlying history model, both non-linear [5, 17, 23, 24] and linear visual representations [13] have been used to visualize the history. Moreover, a set of operations (e.g. navigate, edit, search, annotate) have been supported to allow users to exploit the recorded history for re-visitation or reuse [1, 7, 11, 13, 15, 17, 21, 24].

The two pieces of work most closely related to our research are the graphical history tool for the Tableau database visualization system [13], and Aruvi, a prototype information visualization system developed for supporting the analytical reasoning process [24]. The Tableau graphical history tool [13] records user actions and visualization states as items that can be bookmarked, annotated, revisited, and exported. It was primarily designed to support re-visitation and communication of individual visualizations. Aruvi [24] captures the visualization states of the analytic process and presents them using a horizontal-vertical tree layout. The granularity of the history tracking was determined by application-specific heuristics (e.g., when the mouse pointer leaves a specific GUI panel). Its goal was to provide a high-level overview of all the exploration paths taken and to allow users to navigate back to any previous visualization state during the current analysis. By contrast, our analytic trail technology captures and allows for bookmarking of an entire analytic process (i.e., not just the final visualization state but all the steps that went into its derivation) and re-visitation of the process at any time. Furthermore, our trails can be edited to facilitate the re-purposing of existing analyses to new analytic tasks.

2.2 Asynchronous Collaborative Visual Analytics

Researchers have studied designs to help users collaborate on visual analysis. Sense.us [14] and Many Eyes [4] are web sites that support asynchronous collaboration across a variety of visualization types through view sharing, discussion, graphical annotation, and social navigation. The grid-based web portal described in [15] allows asynchronous users to view, edit, and extend previous visual exploration sessions conducted by other users. Further design considerations for collaborative visual analytics are discussed in [12] and [3].

Existing techniques mostly focus on collaboration by means of static visualization snapshots (e.g. [4, 14]) or spreadsheets of visualization parameters (e.g. [15]). By contrast, our analytic trail technology allows the whole sequence of visual analytic activities encapsulated in a trail to be shared all together at once, and enables users to dynamically interact with, modify, or extend such a trail.

3 Business User Interviews

To inform the design of the analytic trail technology, we conducted semi-structured interviews with six business analysts, whose daily responsibilities include analyzing data visually to derive insights and make business decisions and/or recommendations. All of the interviewees are considered experts in their respective domains, which include market research for emerging technologies; business unit market analysis;

marketing consultation; strategic planning for sales & distribution; software mergers & acquisitions; and financial performance analysis. All analysts work for large, global enterprises. Although the areas of work are diverse, and the active life span of an analytic task ranges from hours to months, our interviews uncovered several common characteristics in how these business users perform their daily analytic tasks.

All six analysts use Microsoft Excel, especially the charting mechanisms within Excel for their data analytic tasks. Except for one user who took extensive training courses in Excel, the other analysts received little to no formal training and largely rely on self-training. Four analysts use additional internal or commercial business intelligence tools to aggregate the data retrieved from a data warehouse in order to generate data sets of a size manageable by Excel. Half of the interviewees analyze both structured, quantitative data and unstructured, qualitative data obtained from multiple sources, while the other half work only with structured data. Those who work with unstructured data manually add structure to the data by annotating and categorizing the textual content so that they can work with the data in Excel. In all cases, visualizations are used not only for detecting trends and outliers during analysis, but also for communicating the findings and the derived insights to their clients, colleagues, and management chain. Only a small number of visualization metaphors (e.g. bar chart, line graph, pie chart) are commonly used, especially when communicating analysis results, due to their simplicity and ease of being understood by business people.

Below we organize our interview findings as they relate to three key aspects that we focus on for data analysis in business environments: analytic provenance, collaboration, and reuse of pre-existing analyses.

3.1 Analytic Provenance

Analytic provenance refers to a historical record of an analytic process, which may include user analytic activities, the data being explored, as well as the insights uncovered during the analysis. All six business users preserve the insights derived from an analysis by manually associating notes and annotations with visualizations. However, this form of provenance is not always sufficient when the analysis needs to be revisited for various reasons. The majority of the users (four out of six) have a need to revisit an earlier analysis, sometimes conducted weeks or even months before, either to obtain the rationale of a decision/recommendation made based on the analysis, to refresh the analysis with updated data, or to document the steps taken in obtaining the final results. To recall the process of an earlier analysis, these users rely on their own memory or brief notes, which are often unreliable, especially for analytic processes that are "*explorative*," "*iterative*," and use "*a trial-and-error approach*." As a result, the analysts often have to manually redo the whole analysis. For these analysts re-visitation of prior analyses is a fairly common business activity, and the tools they are using do not provide sufficient support for this task.

3.2 Collaboration

All of the business users we interviewed team up with other people to perform data analysis for internal or external clients. As a result, they have a constant need for communicating their work with their colleagues. Because many of their colleagues work in different geographic locations and time zones, the collaboration is mostly asynchronous. They currently rely primarily on e-mail, for sending around copies of notes, tables, spreadsheets and in some cases PowerPoint presentations. One user explicitly expressed the desire for a tool to help him more easily share findings of visual analyses with his colleagues. He considered it "*a wasteful business process*" to "*export snapshots*, *cut and paste screenshots of Excel dashboard*," and felt that "*some sort of collaborative tool would be helpful for discussion of data during staff-to-staff interaction*."

3.3 Reuse of Analyses

Although each analytic task conducted by an analyst is usually with new/different data, the data sets often share many similarities. For example, the data sets with marketing information of different countries are likely to have a common list of marketing channels such as television, radio, newspaper, and magazine, as well as similar metrics to measure the effectiveness of the marketing activities through these channels. Similarly, for mergers & acquisitions, the data sets usually contain a common list of entities for company profiles, such as business focus, customers, partners, revenue, size, etc. As a result, the users can often transfer what they have learned and used in previous analyses to a new analysis. However, due to the lack of "replay" support that is easy to customize and use without requiring programming skills, the users often have to manually repeat each of the analytic steps they would like to reuse. When asked about the most difficult, tedious, frustrating, or unpleasant part of their work, five of the six analysts mentioned that they didn't like spending time conducting steps that were almost (but not exactly) the same at different times or across different data sets. An example mentioned was manually mapping the data to the visualization parameters or performing the same type of analysis for different companies or multiple geographic areas. In an attempt to address this issue, three users mentioned that they used Excel spreadsheets created for earlier analyses as "templates," and pasted new data on top of old data in the spreadsheets, so that previously defined functions and visual mappings from the data to the visualization parameters could be reused. When asked about why they didn't use Excel macros for their tasks, the analysts pointed to the lack of skills for creating and customizing macros as the main reason. Representative responses include "*I am not very good at writing macros*," "*I wish I had other people create macros for my purposes*, *but unfortunately we don't*."

We also discovered during the interviews that reuse was not restricted to a user's own analyses. One analyst mentioned that she sometimes studied the reports created by other people for their analyses to learn new ways of analyzing data in Excel, and applied them to her own analyses.

4 Smarter Decisions

In this section, we provide a brief overview of *Smarter Decisions*, a visual analytic tool within which the analytic trail technology was implemented. *Smarter Decisions* is an interactive web-based tool for visual analytics designed to help business users derive insights from large collections of both structured and unstructured data. Fig. 1 shows a screenshot of its main user interface for data analysis. The left hand side of the screen is the query panel where users can build a query using select-lists to retrieve the data (Fig. 1a). The middle portion of the screen is the visualization canvas where the retrieved data is visualized (Fig. 1b). Users explore and analyze data by issuing ad-hoc queries and interacting with the visualizations of the retrieved data (e.g. panning, sorting, filtering). *Smarter Decisions* currently includes several commonly used visualization metaphors such as bar chart, line graph, scatter plot, table, document list, and tag cloud. Based on the technology described in [8], *Smarter Decisions* assists users by automatically instantiating the data in the most appropriate visualization metaphor given the properties of the data, and provides alternate visualization choices on the right hand side of the screen (Fig. 1c). Users can switch to any of these alternates simply by clicking on them.

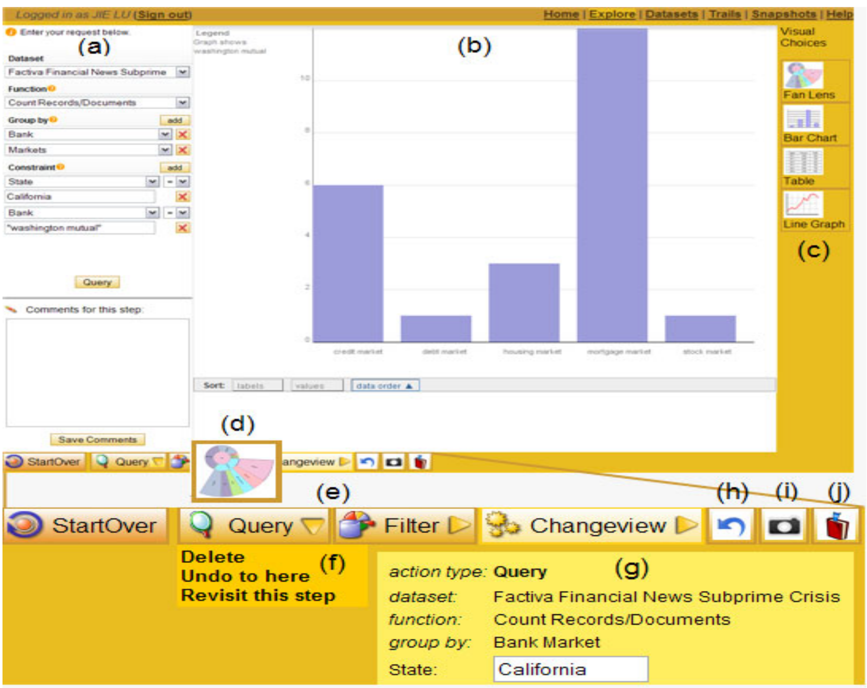


Fig. 1. Smarter Decisions user interface: (a) query panel, (b) visualization canvas, (c) alternate visualization choices, (d) thumbnail of the visualization for a step, (e) trail steps, (f) menu of operations for a step, (g) detail of the action performed during a step, (h) undo, (i) snapshot, (j) bookmark

Smarter Decisions automatically captures the trail of the user actions taken during visual data exploration, such as issuing a query (Query), interacting with the visualization to filter to a subset of the data (Filter), changing to an alternate visualization (Change view), and displays the trail at the bottom of the screen (Fig. 1e, see Section 5.2 for a detailed description). The trail technology is the focus of this paper and is described in detail in the next section. Trails can be bookmarked and restored to replay the actions and data that went into the analytic and decision process, essentially creating a retraceable "memory" of what was done. Trails can be shared to allow for asynchronous collaboration and they can be modified and applied in a new analysis thus facilitating the reuse and/or sharing of an established method for analyzing a given data set. It is our belief that saved trails could also be used to assist with skill ramp-up, when a person is new to the department or organization, or for transfer of expertise when the expert is no longer available.

5 Analytic Trails

The analytic trail technology adopted the trail concept and the semantics-based action taxonomy [10] conceptualized in the HARVEST proof-of-concept system [9]. The design of the trail model, representation, and operations was inspired by prior research as discussed in Section 2, and informed by the interview findings described in Section 3. Compared with HARVEST and other visual analytic tools, the goal of the trail technology in *Smarter Decisions* is to provide an integrated solution to support the needs of visual data analysis in business environments, which include not only personal re-visitation and reuse, but also decision auditing, remote team collaboration, and expertise transfer. In this section, we describe the design considerations for building the trail model, its GUI representation, and the operations it supports to achieve the above goal.

5.1 Trail Model

The trail model defines the representation and organization of analytic provenance. To increase the transparency of provenance, user activities need to be recorded at a semantically meaningful level, easily understood by users. Because *Smarter Decisions* was developed as a visual analytic tool that could be used by average business users in different data domains, we adopted the semantics-based model proposed in [10] to capture the analytic process at a semantic level without using domain-specific heuristics. Low-level user interaction events such as clicks and drags are mapped to a set of semantic but generic user *actions* such as Query and Filter (see [10] for details), which are used as semantic building blocks for the trail. Each action includes a set of parameters to encode the information needed by the system to perform the action, such as the data set, data concepts/attributes, and data constraints. The system also dynamically maintains a summary of the user's task context, which is computed by aggregating the parameters of all the previous actions taken in the user's current line of inquiry. This summary provides the contextual information needed to execute the next action and transform the analytic process from one state to another.

A linear, logical sequence of user actions constitutes an *analytic trail*. A *trail graph* interconnects multiple trails to reflect a non-linear, progressive visual analysis workflow. Trails are connected when the user returns to an earlier state of a trail and creates a new branch of analysis from this state to result in another trail.

Based on the interview finding that the users often annotate visualizations for insight provenance, we added to the trail model a feature that allows text annotations to be associated with individual visualizations created during the analysis. The trail model is also equipped with access control to provide users the flexibility of making their analyses private/public, or sharing a trail with a group of collaborators, which enables the tool to support both personal and collaborative analytic tasks.

5.2 Trail Representation

A trail is represented as a linear sequence of steps. We decided to use a representation of the *active trail* to expose the trail model through the user interface. The active trail includes the sequence of user actions performed by a user during his/her current investigational thread. Bookmarked trails from earlier investigational threads are accessible from a trail library. The decision to only display the active trail was based on two main considerations. First, exposing the whole graph structure of the trail model would occupy too much screen space, distracting users from their primary visual analytic task. In contrast, a linear representation is compact and less obtrusive. Second, a graph-based display increases the complexity and difficulty of trail presentation and interaction, which may cause user confusion about how to interpret the display and how to interact with it. By comparison, a linear display is simpler and easier to understand, which reduces training time and potential cognitive burden on users when using the tool.

Each step in the trail consists of a semantic user action defined in the trail model (e.g. Query, Filter), and the visualization displayed as a result of the action. Information about each step is displayed on two levels. At the higher level, a step is depicted as a button with an icon and a text label to indicate the type of action performed at that step (Fig. 1e). Hovering the mouse over a step shows a small thumbnail of the associated visualization for the step (Fig. 1d). This high-level display enables users to quickly obtain an "at-a-glance" pictorial summary of the analytic process. At a more detailed level, clicking on a step reveals information about the action performed during the step in the form of parameter name-value pairs (Fig. 1g), and a menu of the operations that can be performed in this step (Fig. 1f, which we describe in the next section). The reason to use parameters instead of a natural language-style summary for describing a user action is two-fold: 1) to avoid any misinterpretation caused by ambiguities in natural language, and 2) to allow users to more easily modify the action to reuse its logic within a new context.

5.3 Trail Operations

Smarter Decisions supports operations both at the trail level and at the level of individual trail steps. At the trail level, users can click the bookmark button located at the bottom right of the interface (Fig. 1j) to save the sequence of actions included in the current trail. A unique URL is assigned to each bookmarked trail. Clicking on the URL (e.g., within an email, a blog, or the trail library) results in the trail being restored within the *Smarter Decisions* interface, which replaces any existing trail display at the interface. Once a trail is restored, it becomes the current active trail, which means that it is fully interactive and can also be extended with new user actions to continue the analysis. This mechanism enables users to work collaboratively on an analysis and to adapt an existing trail for new analysis. By default, a bookmarked trail is private so that only its creator is able to access and restore it. A user can change a bookmarked trail s/he created from private to public to allow any other user to access it. Alternatively a bookmarked trail can be shared with specific users, identified by name. At any time, the creator of a trail can delete it from the trail library.

At the level of individual trail steps, users can perform operations such as removing the step from the active trail ("Delete"), removing all the subsequent steps of this step from the active trail ("Undo to here"), and revisiting the step ("Revisit this step") by selecting from the menu associated with each trail step (Fig. 1f). Single and multiple step deletions enable users to remove unwanted actions, especially those performed during the exploration phase of an analysis, and to keep a record of only the sequence that leads to the desired analysis outcome. For convenience, a single-step undo button is displayed at the end of the representation of the active trail (Fig. 1h). Clicking on it results in the last step being deleted. Revisiting a step restores the application to the state that was reached as a result of the user action recorded in that step. The restored information includes an aggregation of the parameters of all the actions performed up to this point of the trail, the visualization displayed, and any user-provided annotations associated with the visualization. Step re-visitation provides a mechanism for users to quickly examine an earlier state of an analysis for understanding the logic and reviewing the result. If a new user action is performed as part of revisiting a previous step (e.g., the parameters are changed, or a filter is applied to the visualized data), a new investigational thread is started, which creates a new trail in the trail graph by branching out from the current active trail. Then the newly created trail becomes the new active trail. This mechanism provides users with the capability of reusing the same steps of a recorded analytic trail for a new analysis without having to manually repeat them. For example, if the user who performed the analysis shown in Fig. 1 wants to investigate the mortgage market as discussed in the articles that mention the state of California, he can reuse the query step in the existing analysis by revisiting this step and execute a filter to just the mortgage market.

The parameter values of categorical, numerical, or keyword constraints for any trail step are editable (Fig. 1g), allowing users to apply the corresponding user action in a different but related context for new analysis. This functionality was designed to enable an analysis to be adapted for use in similar but different analytic tasks. For example, the user can adapt the analysis shown in Fig. 1 for a new investigation in another state, e.g., Texas, by changing the value of the constraint (Fig. 1g) from "California" to "Texas."

Finally, a snapshot button (Fig. 1i), located at the bottom right of the interface next to the bookmark button (Fig. 1j), enables users to export the current visualization to an image that can be embedded in reports and presentations.

6 User Study

We conducted a user study with two primary goals: 1) to evaluate the quality of the support provided by the trail technology with regard to our three focus areas of analytic provenance, asynchronous collaboration, and reuse of analyses, and 2) to gather user feedback on how the design could be improved to better assist users with their visual analytic tasks.

6.1 Study Design

We evaluated the analytic trail technology in the context of the *Smarter Decisions* visual analytic tool. The objective of the user study was *not* to see if use of *Smarter Decisions* with analytic trails was better or faster than a baseline condition, but to examine to what extent the system met (or failed to meet) the needs of business users (e.g. provenance, collaboration, and reuse) and to what degree the features of trails were discoverable by users with little to no prior training. We also felt that a set of baseline metrics would have been difficult to generalize since most of this user population use Excel as a tool, and have no current equivalent to trails.

The data set used for the study was created from the InfoVis publication data [6], which was chosen because it is publicly available and the concepts in this data set are easily understood by the users that would participate in the study. It contains the metadata from 614 papers published between 1974 and 2004, including the title, authors, abstract, topic, references, length, source, and year for each paper.

There were a total of four tasks in the study. The first two tasks were designed to evaluate the tool's support for analytic provenance by asking the users to validate a set of statements based on two existing trails, one for each task. The statements were about research topics, authors, and citations of the InfoVis publications contained in the data set. For example, one of the statements for task 1 was "There are four researchers in total who have published papers on the topic of 'visualizing large data sets' based on this data set," and a statement from task 2 was "The paper titled 'The information visualizer, an information workspace' is the most cited paper in the data set." Task 1 also contained false statements, while all of the statements provided for task 2 were true. Each trail consisted of five or six steps. The users were encouraged to inspect the trails, but were not allowed to extend or modify them.

Task 3 was designed to evaluate the tool's support for cases where multiple people collaborate asynchronously to complete an analysis. Specifically, the users were asked to complete a partially finished analysis by extending an analytic trail which had been started by an imaginary colleague. This colleague had selected a topic for analysis ("dynamic queries"), identified the two researchers with the most papers on this topic, and the papers that each of them published on this topic. The study participants were asked to continue the analysis in order to determine the number of papers each of these two researchers had published across all topics, as well as naming two other topics each researcher had been working on.

Finally, task 4 focused on analysis reuse. For this task participants were asked to conduct a new analysis by either reusing and modifying any trail from the trail library, or starting the analysis from scratch. More specifically, participants were asked to conduct an analysis of the papers published in 1996 to find answers to questions about papers on the topic of "internet" and with the keyword "representation" in their titles or abstracts. A trail that recorded an analysis of the papers published in 1995 with different topic and keyword constraints was included in the trail library, along with trails used for the other tasks of the study.

The trails used for the tasks mostly consisted of Query and Filter steps. The Query steps were used to obtain aggregated (e.g. count) or detailed (e.g. topic, title, author) information of the InfoVis publication data given specific constraints (e.g. a particular topic or year of publication). The Filter steps were used to apply additional constraints to drill down to a subset of the requested data.

It should be noted that the tasks described above didn't cover all forms of use cases for visual data analysis, in particular the synthesis case that brings together multiple threads of analysis conducted by multiple users. Such a use case is often limited to deep analysis conducted by government agencies or scientists. Business users have (relatively) simpler questions that need answering. We believe that the use cases in the study (e.g., auditing an earlier analysis, continuing an analysis started by a colleague, finding an existing analysis from the library to reuse) are representative of the visual data analysis tasks conducted in business environments.

Prior to the main evaluation, we tested the design of the study including the tutorial and tasks with two users and made revisions accordingly. For the main study, we recruited twenty-one users from a large corporate research firm, sixteen of whom were male and five were female. Their ages varied from mid 20s to early 50s, with an average age of mid 30s. All of the users had some experience with general tools that include visualizations (e.g. Google Maps, PowerPoint charts), but were not visualization experts. Before performing the study tasks, each user was given a tutorial on the *Smarter Decisions* tool, during which s/he was instructed to interact with a trail made available for training purposes. All participants received identical training, with the training material being read by the experimenter to ensure conformity. After the training session the participants were given time to ask as many questions as they wanted to ensure understanding. For each task in the study, the users were given the task description and the questions they needed to answer in print form. Access to the trail library containing the trails needed for the tasks was included as part of the user interface for the tool.

We recorded the task completion time for each task, which was counted from the time when the user started interacting with *Smarter Decisions* to the moment when s/he indicated that s/he had finished answering all questions for the task. We collected subjective feedback from the users through questionnaires at the end of every task. The questionnaires included a set of questions for which the answers were measured using a Likert scale that ranged from 1 for "strongly disagree" to 7 for "strongly agree" (see Table 1), and open-ended questions such as "what was difficult about using trails" to further collect user comments and suggestions.

6.2 Study Results

In this section we report on objective and subjective data collected from the study. Section 7 provides further discussion of the study findings and their implications.

The average training session across all the users lasted 8.89 minutes. 81% of the users spent less than 10 minutes on training (the variable being the number of follow-up questions each participant asked). One user required over 20 minutes of training, in which he asked many detailed questions while exploring various components of the interface. By comparison, the other users asked fewer questions and largely familiarized themselves with the interface during task performance.

In spite of this very brief training time, all participants were able to complete their tasks. The longest study completion time was 40.07 minutes, which was about 10 minutes per task. We observed during the study that longer training mostly yielded shorter total task completion times. Fig. 2 depicts the relation between training time and total completion time for each participant. The correlation coefficient between the two time variables is –0.50 with a *p*-value of 0.02 (–0.67, *p*-value < 0.001 if excluding the participant mentioned above with the longest training time), indicating a weak negative linear correlation.

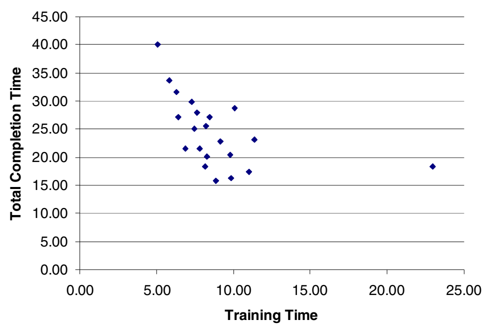


Fig. 2. The relations between the training time and the total completion time (in minutes)

Among the four tasks, task 1 and task 2 required similar completion times (mean 5.79, std. dev. 2.20 vs. mean 5.21, std. dev. 1.75 in minutes for tasks 1 and 2 respectively, with no statistically significant difference between them) since they were similar to each other by design (i.e., validating three statements about paper authors or citations based on an existing trail). Task 3 took the longest time to complete (mean 9.48, std. dev. 2.57 in minutes). This was because in comparison with the other tasks, each user needed to perform more analytic steps and answer more questions in order to complete task 3. Task 4 contained the least number of questions, yielding the shortest completion time (mean 3.91, std. dev. 1.87 in minutes).

Fig. 3 shows the subjective data measured using a Likert scale. The results indicate that the users on average felt positive about being able to easily understand and use the concept of analytic trails and associated steps for quickly understanding and validating analysis results (Q1-Q4). Two users explicitly pointed out that they liked being able to validate analysis results by revisiting the trail saved for the analysis so they didn't need to start from scratch. 71% of the users responded positively that trails would improve the transparency of decision making if used in their teams or in their company (Q5). Also 19% of the users felt that they could see the potential usefulness of trails in increasing the transparency of decision making, but since they didn't do visual data analysis as part of their current assignment for the work, they chose to remain neutral on this subject.

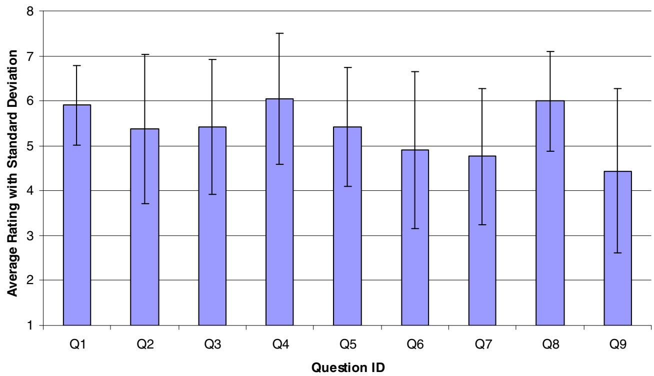


Fig. 3. The average ratings (with +/– 1 std. dev. as error bars) for the Likert-scale questions (1: strongly disagree; 2: disagree; 3: somewhat disagree; 4: undecided; 5: somewhat agree; 6: agree; 7: strongly agree)

Table 1. The Likert-scale questions included in the questionnaires following the tasks

ID	Question
Q1	Easy to understand the concept of trail and associated steps
Q2	Easy to understand a previous analysis based on its trail
Q3	Easy to validate analysis results by revisiting trail steps
Q4	Faster to validate analysis results using trails than using other sources
Q5	Trails would improve the transparency of decision making
Q6	Easy to determine how to extend an trail to complete the analysis
Q7	Helpful to be able to extend a saved trail to complete an analysis
Q8	Helpful to adapt an existing trail and re-apply it to a new analysis
Q9	Easy to find a trail most relevant to task at hand in the library

Compared with the almost universally positive opinions about the usefulness of trails for understanding the provenance of existing analyses (task 1 and task 2), the users were less enthusiastic about extending a saved trail to complete an analysis started by someone else (task 3). Only 52% of the users agreed or strongly agreed that it was easy for them to determine how to extend a trail to complete the analysis (Q6). Three users who expressed negative opinions on this subject wanted to distinguish their new analytic steps from existing ones, but the current design lacked support for this feature. Only 38% of the users agreed or strongly agreed that they found it helpful to be able to extend a saved trail to complete an analysis (Q7). Four users commented that they preferred to follow their own logic and thought process to perform an analysis, and didn't like to start from a trail created by somebody else.

For task 4, one third of the users chose to start from scratch rather than reusing a trail from the trail library, citing the relative ease of performing a new analysis to complete the task, and the estimated degree of difficulty in finding a relevant trail from the trail library as the two main reasons for such a decision. Among the participants who selected a trail from the trail library to reuse and adapt it to complete the required analysis, on average they agreed that it was helpful to adapt an existing trail and re-apply it to a new analysis (Q8). However, only 43% of them successfully found the most relevant trail for the task from the trail library, and agreed or strongly agreed that they could easily find a trail in the library most relevant to their task at hand (Q9). The other 57% reused a sub-optimal trail instead.

For the open-ended questions, when asked about what was easy about using trails, the users mentioned that the high-level information of a trail was easy to understand, the details of trail steps were easy to access and understand, and a trail was easy to navigate. These aspects made it easy to revisit and follow the logical path of an analysis and validate the results along the trail. Regarding what was hard about using trails, there were three common complaints. First, seven users felt that the high-level information about trail steps (i.e., descriptions of action types in the step buttons and small thumbnails of visualizations when the step buttons are moused over, as illustrated in Fig. 1d-e) was too abstract and desired more details at a glance. Second, seven users expected the visualization associated with a step to be shown by clicking on the button of this step instead of being required to select the "Revisit this step" option from the step's menu. Third, because the current design of trail representation only displays one active trail at a time, three users felt that it would be difficult to use trails for analyses involving multiple active investigational threads in parallel.

7 Discussion

We developed the analytic trail technology for our *Smarter Decisions* visual analytic tool with the goal of increasing the transparency of analytic provenance as well as supporting asynchronous collaboration and reuse of visual data analyses. During the user study, we received valuable feedback on the effectiveness of our development and how the technology could be further improved to achieve this goal. Here we discuss the feedback and the remaining challenges.

7.1 Analytic Provenance

Study results indicate that the users were receptive of the analytic trail technology and positive about the value of trails for increasing the transparency of analytic provenance. Trails were shown to be effective in helping the users understand the logic of existing analyses as needed to validate the results/statements generated. However, the feedback from the users suggests that the current design of trail GUI representation could be improved to strengthen the benefits of trails for analytic provenance. Particularly, several users commented that the design of the representation at the trail level needs to provide an effective summary for them to quickly understand the analytic process. In some cases, using action type descriptions (e.g. Query, Filter) and small thumbnails of visualizations to describe trail steps was not sufficiently informative for the users to quickly understand the analytic process that was undertaken. The users were required to perform multiple mouse clicks to get to the details of the trail steps one step at a time, and rely on their memory to piece together the logic of the analytic process. In other cases, multiple semantic actions might correspond to one logical action in a user's thought process, but individual steps displayed in a linear sequence didn't reflect the logical relation or groupings of the steps. As a result, for a long trail with a large number of steps, the trail representation became too low-level, making it difficult for the users to grasp the logical flow of the analysis. It is a challenge to find a single level of granularity for trail representation that works well in all the cases, especially without the help of domain heuristics.

A better solution may be to dynamically adjust the granularity of trail representation based on the characteristics of the analysis. For example, when the analysis includes a small number of steps with simple logic, details of the steps can be made visible at the trail level. For a complex analysis with many steps, individual steps that correspond to one logical construct of the analysis can be "chunked" together. Such logical chunking can be performed manually by the user who conducts the analysis, or automatically by the system based on machine learning and mining from user analytic behaviors. User-provided descriptions can be associated with each chunk to improve understanding. This solution requires the trail model to be augmented to support hierarchical organization of actions, and interaction mechanisms to support multi-level zoom in/out for the trail display at the interface.

In addition to determining the right level of granularity intelligently for trail representation, the trail technology should also support more intuitive interaction mechanisms for trail step re-visitation (e.g., clicking on the button of a step to revisit instead of selecting the revisit option from the step's menu).

7.2 Collaboration

Interestingly, the use of trails in asynchronous collaboration around visual data analysis was not as well-received as we had hoped. The log and questionnaires from the study revealed two primary explanations for the relatively lukewarm response to this feature. First, the participants were not sufficiently motivated to understand all the details of an analysis conducted by their (imaginary) collaborator in order to complete the required task (task 3). This could be due to the fact that the task was conducted in the laboratory setting with an imaginary collaborator instead of the users' actual working environment with real persons as collaborators. Some users felt that they only needed to know the outcome of an analysis and didn't really care about how the analysis was conducted. Similarly, these users only felt compelled to share their analysis results but not the process. One user mentioned that he was not sure if he would ever share the trail of his analysis with others since his collaboration with others were loosely coupled, for which sharing analysis results would be sufficient. Therefore, he would want to save the trail of his analysis for personal use, but not for collaboration. However, the current design of the trail technology makes it difficult for the users to obtain or share the information about "what" (analysis results) without sharing the detail about "how" (analytic process). Second, some users didn't want to follow other people's thought process in order to perform an analysis collaboratively. They preferred following their own logic and didn't want to mix the record of their analytic steps with the record of the collaborators'. For these participants, trails were viewed as more valuable for recording, navigating, and adapting an analytic process for personal use, rather than for communicating with other people as part of a collaboration.

The above findings suggest that the most appropriate structure and granularity of trail representation depend on not only the characteristics of the analysis (e.g. complexity) as discussed in the previous section, but also the purpose for analytic provenance. For example, representation at the level of individual analytic steps may work well for personal use, which includes viewing the details of the analysis, but a level based on logical grouping, or authorship of trail steps, may be more appropriate for loosely-coupled collaboration. Also if the goal of the review of individual steps is to audit the decision, or to increase the transparency of what data was included in a decision, then the details could be made available on demand. Furthermore, the trail representation should make it easy to navigate between multiple related trails created by different users as part of a collaborative analysis.

7.3 Reuse of Analyses

We designed task 4 of the user study to focus on evaluating the effectiveness of trails with regard to facilitating reuse of analyses. We were surprised that as many as seven users didn't even try to browse the trail library to find a trail they could reuse for new analysis. When asked about why they didn't reuse a trail, some users said that they felt they could perform the task easily from scratch without the need to reuse an existing trail. More complex tasks might provide greater incentive for the users to reuse trails of existing analyses, and longer-term use of our tool could reveal greater benefits of trail reuse, which we weren't able to test in the laboratory setting where we limited the complexity of the tasks so as not to overwhelm the study participants, all of whom were first-time users of our tool. Eight users didn't correctly identify the most relevant trail for the task when browsing the small library of six trails. Instead of selecting a well-suited trail we had placed in the library as a public trail, they chose a trail they had used in one of the previous tasks. These users felt that without built-in search support for the trail library, it was difficult to determine the relevance of a trail for a new task based on its brief description in the library, and too time-consuming to examine all the trails to find the most relevant one, especially for trails created by others. They expected this problem to become more serious as the trail library grew. This study result points to the need to provide rich search support for trails in the library (by author, by keyword, by date, etc.), or trail reuse will likely be limited to cases where the users know in advance which trail to reuse.

We also observed that the users who reused a trail for task 4 first spent time inspecting the trail to find the steps relevant to the task at hand, then deleted all of the unwanted steps before modifying and adapting the relevant steps to suit their needs. These users expressed a common desire for the support to easily find the relevant parts of the trail and easily manipulate trail steps, such as moving (e.g. by dragging and dropping) a trail step from one position of the trail to another and making a copy of one or more trail steps. Therefore, in addition to finding a relevant trail, effective reuse of analyses requires system support for easily locating the parts of the trail to be reused and adapting them for new analysis. How to design and implement such new features without making the interface overly complex and reducing its usability is a challenging problem, which is next on our research agenda.

To further support reuse of analyses, we also plan to develop new functionality that generalizes trails to create trail templates that can be easily customized and applied by different people on different data and analytic tasks, with the goal of facilitating the use and sharing of best practices and helping more effective skill/expertise transfer.

8 Conclusion

In this paper, we present the analytic trail technology in the context of a visual analytic tool designed to empower business users to derive insights from large amounts of data. Informed by the findings from the interviews with several business users about their visual analytic activities, we present the design of the trail model, its GUI representation, and the operations it supports with the goal of providing a mechanism to increase the transparency of analytic provenance as well as support asynchronous collaboration and reuse of visual data analyses in business environments. To facilitate analytic provenance, the trail model represents user analytic activities with semantic actions (e.g. Query, Filter, Change view) and captures a linear, logical sequence of actions into a trail. Multiple trails of an analysis are organized into a graph-based structure to reflect a non-linear, progressive visual analysis workflow. The active trail which corresponds to the sequence of actions performed during a user's current investigational thread is displayed at the GUI. The trail GUI representation was designed to help users easily navigate a trail and obtain information about the encapsulated user actions and visualization results for understanding the provenance of the analysis. Trail operations such as bookmarking, sharing, revisiting, and editing are provided with the goal of facilitating re-visitation, asynchronous communication, and reuse of analyses.

We designed and conducted a user study to evaluate the effectiveness of the analytic trail technology at supporting its goal of provenance, reuse and collaboration. The results indicate that most participants found trails to be useful for capturing and understanding the provenance of an analysis. However, with the current design, trails were considered to be more valuable in recording, navigating, and adapting an analytic process for personal use, rather than for communicating the analytic process to other people as part of a collaboration. The results also indicate that search support for easily finding relevant trails or relevant parts of a trail is critical to support the goal of adaptation and reuse of analyses. These findings suggest areas where trails provide the greatest value and point out directions for future research in the area of capturing analytic processes.

References

1. Bavoil, L., Callahan, S., Crossno, P., Freire, J., Scheidegger, C., Silva, C., Vo, H.: Vistrails: Enabling interactive multiple-view visualizations. In: IEEE Vis (2005)
2. Brodlie, K., Brankin, L., Banecki, G., Gay, A., Poon, A., Wright, H.: Grasparc – a problem solving environment integrating computation and visualization. In: IEEE Vis (1993)
3. Chinchor, N., Pike, W.: Science of analytical reporting. Information Visualization 8, 286–293 (2009)
4. Danis, C., Viegas, F., Wattenberg, M., Kriss, J.: Your place or mine? visualization as a community component. In: CHI (2008)

5. Derthick, M., Roth, S.: Example based generation of custom data analysis appliances. In: IUI (2001)
6. Fekete, J.-D., Grinstein, G., Plaisant, C.: IEEE InfoVis 2004 contest data set (2004), `http://www.cs.umd.edu/hcil/iv04contest`
7. Goodell, H., Chiang, C., Kelleher, C., Baumann, A., Grinstein, G.: Collecting and harnessing rich session histories. In: IV (2006)
8. Gotz, D., Wen, Z.: Behavior-driven visualization recommendation. In: IUI (2009)
9. Gotz, D., Wen, Z., Lu, J., Kissa, P., Zhou, M., Cao, N., Qian, W., Liu, S.: HARVEST – Visualization and analysis for the masses. In: IEEE InfoVis Poster (2008)
10. Gotz, D., Zhou, M.: Characterizing users' visual analytic activity for insight provenance. In: IEEE VAST (2008)
11. Groth, D., Streefkerk, K.: Provenance and annotation for visual exploration systems. IEEE Trans on Vis. and Comp. Graphics 12(6) (2006)
12. Heer, J., Agrawala, M.: Design considerations for collaborative visual analytics. Information 7(1), 49–62 (2008)
13. Heer, J., Mackinlay, J., Stolte, C., Agrawala, M.: Graphical histories for visualization: Supporting analysis, communication, and evaluation. IEEE Trans. on Vis. and Comp. Graphics 14(6), 1189–1196 (2008)
14. Heer, J., Viegas, F., Wattenberg, M.: Voyagers and voyeurs: Supporting asynchronous collaborative information visualization. In: CHI (2007)
15. Jankun-Kelly, T., Kreylos, O., Ma, K., Hamann, B., Joy, K., Shalf, J., Bethel, E.: Deploying web-based visual exploration tools on the grid. IEEE Computer Graphics and Applications 23(2), 40–50 (2003)
16. Jankun-Kelly, T., Ma, K., Gertz, M.: A model for the visualization exploration process. In: IEEE Vis (2002)
17. Kadivar, N., Chen, V., Dunsmuir, D., Lee, E., Qian, C., Dill, J., Shaw, C., Woodbury, R.: Capturing and supporting the analysis process. In: IEEE VAST (2009)
18. Keim, D., Mansmann, F., Schneidewind, J., Ziegler, H.: Challenges in visual data analysis. In: IEEE InfoVis (2006)
19. Kobsa, A.: An empirical comparison of three commercial information visualization systems. In: IEEE InfoVis (2001)
20. Kreuseler, M., Nocke, T., Schumann, H.: A history mechanism for visual data mining. In: IEEE InfoVis (2004)
21. Kurlander, D., Feiner, S.: Editable graphical histories. In: IEEE Workshop on Visual Language, pp. 127–134 (1998)
22. Lissack, M.: Of chaos and complexity: Managerial insights from a new science. Management Decision 35, 205–218 (1997)
23. Ma, K.: Image graphs – a novel approach to visual data exploration. In: IEEE Vis (1999)
24. Shrinivasan, Y., van Wijk, J.: Supporting the analytical reasoning process in information visualization. In: CHI (2008)
25. Thomas, J., Cook, K. (eds.): Illuminating the path: The research and development agenda for visual analytics. IEEE Press, Los Alamitos (2005)
26. Wong, P., Thomas, J.: Visual analytics. IEEE Computer Graphics and Applications 24, 20–21 (2004)
27. Xiao, L., Gerth, J., Hanrahan, P.: Enhancing visual analysis of network traffic using a knowledge representation. In: IEEE VAST (2007)
28. Yi, J., Kang, Y., Stasko, J., Jacko, J.: Toward a deeper understanding of the role of interaction in information visualization. IEEE Trans on Vis. and Comp. Graphics 13(6) (2007)

Exploration Views: Understanding Dashboard Creation and Customization for Visualization Novices

Micheline Elias[1,2] and Anastasia Bezerianos[1]

[1] MAS Laboratory, Ecole Centrale Paris, France
[2] SAP Research, Business Intelligence Practice, France
{micheline.elias,anastasia.bezerianos}@ecp.fr

Abstract. With the increase of visualization platforms targeting novices, researchers are now focusing on gathering insights regarding novice user practices. We describe the design and evaluation of Exploration Views (EV), a system that allows novice visualization users to easily build and customize Business Intelligence information dashboards. EV provides an intuitive environment for dynamically creating, rearranging, searching and exploring multiple visual data representations from diverse data-sources. These aspects aid users to better retrieve, experiment and familiarize themselves with their data. We evaluated EV with both novice and expert dashboard designers and report here (i) how novice users interact with the system, (ii) differences in how novice and expert users react to a dashboard systems that targets both, and (iii) provide new design guidelines for practitioners building dashboard applications, on the needs of novice visualization users.

Keywords: synchronized views, interface customization, novice users, visual queries, business intelligence dashboards.

1 Introduction

Business Intelligence (BI) deals with the collection of processes and software that supports organizations in understanding large datasets, retrieving and analyzing information and making decisions. The value of visual presentation of data was identified early on in this field and usually takes the form of dashboards, interfaces resembling an automobile's dashboard that organize and present information in an easy to read manner [27]. BI visualization dashboards provide collections of multiple visual components, such as charts, on a single view so that information can be monitored at a glance [10]. Appropriate visual representations in dashboards, using colors, size and shape, are combined with interactive exploration [10] to amplify human cognition and enhance information understanding [1].

The current life-cycle of a BI dashboard involves multiple actors [3], including end-users and business analysts. In user-centered BI [11] end-users intervene and provide feedback to the business analysts that create customized dashboards to meet user needs. This feedback comes at different stages of the dashboard design and setup, and involves a large amount of communication between business analysts and end-

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 274–291, 2011.

users, in order to define functional specifications and a positive user experience. Thus there is an intrinsic delay introduced in any end-user requirement change or customization request. Moreover, current market trends (seen in the Gartner Survey [24]), recognize dashboards and data analysis are one of the driving forces adding value to companies. This indicates a tendency for an expanding end-user population with diverse needs, requiring quick access to customizable dashboard technology, bypassing dashboard designers. This is reflected in advances in easily customizable dashboard visualization systems (e.g. [28,29]), it coincides with advances in easily accessible information visualization environments for novices (such as ManyEyes [31]), and is also reflected in the increasing focus of the InfoVis research community on how to support novice users [12,15]. Nevertheless, there are still many questions about how novice users interact with actual information visualization systems.

To better understand how novice users interact and build visualizations on their own (without dashboard designers or other human mediators) we created Exploration Views (EV). EV is a dashboard prototype (**Fig 1**) build specifically to empower end-users, following guidelines from previous work on visualization novices. To evaluate its effectiveness, we compared how dashboard experts and novices reacted towards EV when creating visualizations for realistic tasks. Finally, through our observations we derived a set of additional design guidelines for dashboards targeting novice users.

Our contribution consists of: (i) putting together all guidelines for novice users (empirical or tested) in a fully functional system that can be tested in practice, (ii) creating a system that also includes advanced functionality required by expert users, and (iii) testing the system under realistic tasks to gain insights.

Fig. 1. The Exploration Views consists of: (a) the main Dashboard and its components. (b) The Clipboard drawer widget that includes templates for charts (top) and data widget samples (bottom) that the user has created as a pallet. The drawer can be closed when not needed. (c) A visual representation of the global data filters that are active at any point. (d) And a set of other functions, such as search, save, share, etc.

2 Related Work

In this section we will present related research in the fields of dashboard creation and visualization construction, especially in the context of novice visualization users.

Dashboards. The majority of the commonly used commercial BI dashboards (e.g. [8,23,26]) assume that from the end-user perspective the dashboard design will remain consistent, and focus more on easy report generation. To create new or customize existing dashboards, end users need to either contact trained dashboard designers, or install separately provided dashboard design components and spend time to train themselves. There are some notable exceptions. Tableau [29] incorporates the design functionality in the dashboard. Spotfire [28] also accommodates novice users with a flexible drag-and-drop interface for customizing visualizations on the fly. Finally, the prototype Choosel environment [4] allows for easy visualization creation and further exploration of the visualized data, but does not yet provide advanced functionality (e.g. hierarchical data representation as in OLAP data models [22], with drill-up/down functions), required in real life business data analysis.

Visualization creation. A variety of visualization toolkits allow the creation of sophisticated visualization charts and combinations of them (e.g. Flare[1], Silverlight[2], JavaScript InfoVis tk[3], ivtk[4]). These do not target novice visualization users, and usually require additional programming to create the visualizations, to synchronize them in dashboard views, and to connect them to data-sources. Several environments, such as Improvise [32] or Snap-Together [21], allow the construction of highly customized, powerful and fully linked visualizations (for a comprehensive list see [25]). Nevertheless their setup cost and learning curve is usually high and they target visualization experts rather than novice users.

Visualization for novices. On the other hand, there is a set of visualization platforms that specifically target novice users. For instance, the Google analytics platform [6] provides a highly customizable dashboard with good default visualization charts for web-related data. Other web-based visualization platforms easily accessible to novice visualization users, like ManyEyes [31], Sense.us [16], or Polstar [5], usually restrict users to a viewing and sharing a single visible chart at a time. Thus creating a dashboard is a laborious process performed outside the platform, and linking of multiple charts is no longer possible.

Novices in visualization. With the increase of visualization platforms targeting novices, researchers have started gathering insights into their practices and behaviors. Heer et al. [15] group different visualization systems by how skilled their users have to be, and note that tools supporting detailed data exploration target mostly experts, whereas it is mostly communication/reporting systems that accommodate novices. They also suggest that visualization tools for novices should: allow user-friendly data input for common data formats; automatically provide visualizations or reasonable

[1] http://flare.prefuse.org/
[2] http://www.silverlight.net/
[3] http://thejit.org/
[4] http://ivtk.sourceforge.net

defaults based on data types; and use contextual information to clarify the displayed data and encoding. Grammel et al. [12] investigate how novices create single charts, using a human mediator to interact with the creation interface. They found that the main activities performed by novices (apart from communicating their specifications to the mediator) are data attribute selections, visual template selections, and viewing and refinements of the above. They observed that subjects often had partial specifications for their desired visualization, and faced three major barriers: selecting correct data attributes, choosing appropriate visual mappings, and interpreting the visual results. They also verified suggestions from [15], and provided additional guidelines, such as facilitating searching of attributes, automatically creating visualizations when possible, providing explanations, and promoting learning.

Our work extends the above in the following ways: Using guidelines from [12,15], we build a fully functional dashboard visualization system, and observe how novice users use the interface to create dashboards (without a human mediator [12]). We compare our observations of novices to reactions from experts, and derive additional design guidelines for visualization dashboards that target both user groups.

3 Exploration Views (EV) Prototype

In this section we describe the design rationale behind Exploration Views (EV) and the functionality that supports novice users, based on guidelines from previous work. We will first explain the main goals of EV: easy creation, customization and interaction with multiple visual representations in a unified dashboard environment.

Although dashboard users may have clear questions regarding their data, they can be novices when it comes to visualization, and naive as to how dashboard software can support their needs. An *easy creation* process is essential for novices to experiment with dashboard designs fast. As seen in [12], visualization novices often have partial mental specifications for their visualization needs and tend to refine and change their designs. To ensure a user-friendly dashboard creation, the sequence of steps needs to be simple, with continuous visual feedback on the final outcome of user choices. Moreover, as novice users may have no previous knowledge of what visual templates and representations are possible for different data types, EV must provide chart suggestions and templates to choose from (also suggested by [12,15]). Finally, EV should support common data formats (recommendation in [15]), and users should not be restricted by data storage architecture, but be able to use data from multiple data-sources (e.g. personal data in excel and OLAP data from their enterprise).

As novice users create dashboards, they may need to explore and try out alternative visual templates and representations to learn what meets their needs. Requirements and tasks can also change over time, requiring different data representations and data filtering mechanisms, or new data sources. EV should support iterative visualization specifications (suggested in [12]) by being *easily customizable and adaptable*.

It is important for novices to see the effect of their choices immediately, promoting exploration and experimentation with visualizations. Thus in EV customizations *become active immediately,* ensuring that customization is tightly coupled with the *visual analytics process* (recommended by [12]). EV further supports the visual analytics process with functions such as saving and sharing, text search mechanisms, fully linked visualizations and visual queries, and other data exploration mechanisms.

Finally, if the design of EV supports both low-cost experimentation with visual templates and visual analytics processes, it can help novice users to become more accustomed to visualization creation and analysis, *promoting learning*.

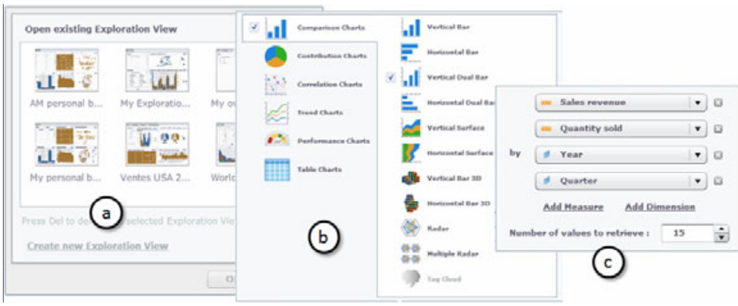


Fig. 2. (a) Creation dialogs with visual examples of templates or saved dashboards. (b) Chart creation dialog with recommendations of charts appropriate to data categories and analysis types, sorted from simpler to more complex. (c) Data Category and measures selection.

3.1 Easy Creation

When creating a new dashboard the system provides miniature icons of saved dashboards and several template dashboards to choose from (**Fig 2.a**). The visual representation of possible dashboards helps users recall designs (as opposed to remembering dashboards solely based on names [20]), and their titles are descriptive so novices can start from a reasonable existing dashboard design, minimizing new content creation. These dashboard templates can be also shared with others.

If none of the existing dashboard templates meets the user needs, she can choose to create a dashboard from scratch. The user is guided by the system by first asking her to select data sources of various format types that are supported. When the data sources have been selected, the users can populate their dashboards with charts.

The dashboard work area (**Fig 1.a**), acts as both the design space for the dashboard, as well as the active dashboard when content is added. Here users can add charts, tables, and data filters. Possible components to add can be found in the Clipboard drawer (**Fig 1.b**), which can be closed when creation (or customization) ends. Components from the Clipboard can be dragged-and-dropped in the work area, and EV automatically realigns existing components to accommodate new additions.

When adding charts (i.e. visual templates), users need to select data attributes (e.g. sales, year of sale, etc), and match them to visual mapping (a visual property in the template). This is a challenging task for novice users [12] and EV aids them (as in [9,35]) by providing reasonable mapping and visual template recommendations (**Fig 2.b**). After selecting one or more data attributes in the form of measures (data attribute, e.g. sales prices **Fig 2.c**) and categories (possible groupings of the attribute, e.g. per state, year), the system presents appropriate analysis types for this data. For example for a dataset of sales per state over the period of 2 years, possible analysis types include comparisons (sales by state or year), contributions (the percentage of sales per state over all sales), or trends (the evolution of sales over time). If the user selects one of these analysis types, the system automatically recommends visual templates (charts) suitable to the specific analysis and to the nature of the attributes (e.g. how many measures and categories are selected, if they are categorical, ordinal, etc), using a default visual mapping. These recommendations are ranked from simple and more commonly used charts to more complex ones, and aid novices to explore reasonable alternatives. Available visual templates include charts, gauges and tables.

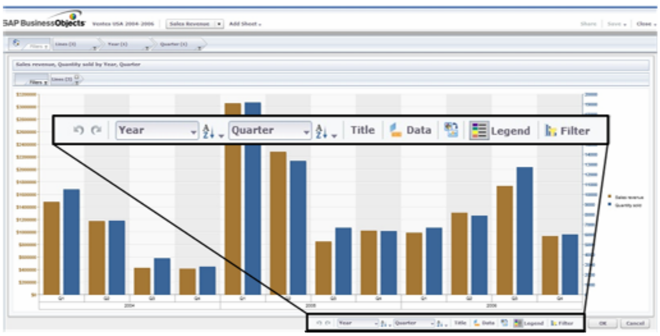


Fig. 3. The "Explore" mode is a detailed view of a chart, where users can undo/redo actions, choose data categories, sort items, change data sources, switch chart templates, and create local data filters. To avoid switching to the "explore" mode, part of this functionality is also available as a drawer toolbar for each chart in the main dashboard, with added options to lock the chart.

3.2 Ease of Customization

During the life of a dashboard, user needs may change; new hypothesis may need to be explored and new data sources to be added. Especially given novice users' tendency for iterative visualization specifications [12], EV has several features for user friendly customization, such as actions to add/remove components, change their properties, duplicate their content, and explore sub-parts of the data using data filters.

Users can add new components on the dashboard or drag-and-drop existing ones from the Clipboard. If the components are visualization templates, the user has to go through data, attribute and mapping selection as described before. If the components are saved in the Clipboard, we use the properties of the component as it was stored.

Existing components on the dashboard can be refined and customized by clicking the "explore" icon of the component. This provides a detailed view of the component (**Fig 3**), together with choices for datasets, visual templates, attributes and mappings (as described in the **Sec 3.1**). Changes are immediately reflected on the component in an automatic preview, so that users can experiment with different choices [12].

Users can also customize the dashboard layout itself through functionality for hiding/showing dashboard components, dragging to rearrange them, resizing them, etc. EV intelligently resizes components to fit in the dashboard (e.g. **Fig 4**).

At any point users can save content for later use (in this or other dashboards), by dragging them in the Clipboard Drawer. This allows exploration of alternatives without risking loosing existing components. To further ensure that novice users are not afraid to experiment with their dashboards, all actions are reversible through undo history functionality applied on the entire dashboard and locally on specific charts.

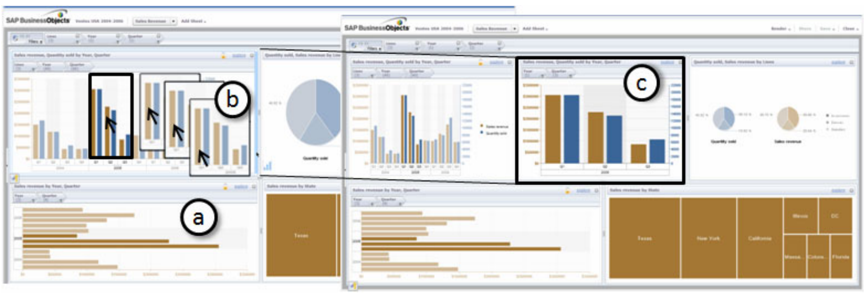


Fig. 4. (a) Users select subset of their data, that get highlighted in all charts. (b) They drag their selections outside the chart to create a cut-out clone visualization (c), of the same type (here bar chart) with the selected data only. A blue highlighted bar (b) indicates the new chart will be placed on the right of the original, with existing charts shift to accommodate it (c).

3.3 Supporting Visual Analytics Process

Apart from the creation and customization functionality, EV provides support a series of functions that aid the visual analytics process. EV has text search support that follows the guidelines of dynamic queries [33]: as the user searches text the corresponding text components on the dashboard that match it (axis labels, categories, tags and any other textual information) get dynamically highlighted.

All our visual representations are essentially linked coordinated views [25,32] of the underlying datasets (unless users give them local filters, discussed later). We thus extend the dynamic query functionality to any visual selection or query in the dashboard: if the user selects in one chart bars representing the 3 quarters of a year, these data instances are also highlighted on all other charts (**Fig 4.a**).

EV further encourages users to explore their data by providing global and local data filters, i.e. constraints on the data entries displayed. Filters can affect the entire dashboard (global), or specific components (local). Thus users can explore "what-if" questions without changing completely their dashboard. Our filters focus on data dimensions: enterprise data usually follow the multidimensional OLAP model [22], where data is organized in hierarchical categorical or ordinal dimensions (e.g. sales grouped per country and state, per year or month, etc). Each dimension has a set of values, which we call categories. Users can drag-and-drop Data-Filter components from the templates (**Fig 5.a,b**), that can be used to restrict data categories.

To enable users to dynamically explore filtered data and compare categories, EV provides faceted navigation components [13,14]. Data-Filter-Controller components are visual components, which get attached to a data dimension, similarly to filters. But instead of constraining the dimension categories, they provide a visual list of the categories (**Fig 5.d**) that users can toggle on/off. Users can easily create and dismiss controllers to ensure that only the desired ones are visible at any given time, reducing clutter in the faceted navigation space (a challenge of faceted navigation [14]).

During data analysis and exploration, users may want to clone and explore in parallel interesting data. EV allows visualization cloning by using the Clipboard (full cloning) or by dragging and cloning parts of the data. For example in the bar chart in **Fig 4**, the user wants to explore in detail sales for the 3 quarters of 2005. By selecting and dragging only these bars, she creates a cut-out clone visualization of just the selected data, which keeps the initial component properties (dataset, visual template, categories, mappings). She can thus explore data subsets without creating components from scratch. This cloning process is essentially a fast local filter creation.

If the underlying data-sources follow the classic OLAP data organization, users have access to the traditional OLAP functionality [22], such as sorting, drill up/down, pivot, roll-up, etc. Users can also flip the chart axis to experiment with different chart layouts. This exploration functionality is necessary for expert users. It is present in the "explore" chart mode, but also in the main dashboard in the form of a widget drawer for each chart to avoid opening the "explore" view (e.g. in [29]). At any point users can undo/redo their actions on each chart. Finally EV provides functions to support long term and collaborative visual analysis, such as save and share.

To help users follow visual changes during data exploration, when using OLAP functions or filters, changes are smoothly animated to ensure visual momentum [34].

Fig. 5. Filters and Faceted navigation creation. (a) The user can drag-and-drop a new global filter and choose the data categories that will be visible on the dashboard. The filter is either directly applied, or can become a filter control component for faceted navigation (b). All active global filters are visible at the top of the dashboard (c), and local filters on the top of charts (d).

4 Exploration Views Technical Platform

EV is not just a lab prototype, but a fully operational BI dashboard that can handle large amounts of data from different sources. Although this paper focuses on the interaction aspect of EV, we briefly describe its architecture. EV has 3 layers. The first is a web client based on the Adobe Flex that contains the dashboard layout, the exploration and faceted navigation components, and the chart manager. The chart manager CVOM (Common Visualization Object Model) provides a set of services for chart creation and recommendation based on the analysis users want to perform and data measures. The second layer is a server containing a user session manager (as in [5]), a dashboard state management, and a component that synchronizes different data sources to enable the multi-chart visualizations. The third layer consists of an in-memory database for quick data storage and retrieval [30], and a semantic layer that helps us to connect visual queries from multiple charts and have synchronized views.

5 Study Design

In this section, we define the purpose of our study, its design and methodology.

Our goal is to examine how novice visualization users create dashboards. We attempt to answer three questions: (i) is it easy for novices to discover visualization and dashboard functionality in an interface specifically designed using previous guidelines for novices; (ii) are the existing guidelines effective and what is missing from the literature, especially in the context of dashboard visualizations; and (iii) do novice and expert visualization users react similarly to such an interface.

Q1. Can novice users discover functionality and build dashboards? We created EV following guidelines by [12,15], and asked users to construct visualizations to answer questions, without the help of a human mediator (contrary to [12]). We wanted to see if novices can discover functionality, learn the interface and answer questions based on the created visualizations. Thus we provided no interface training and observed difficulties in interface discovery and visualization creation process.

Q2. Are the existing guidelines enough? Our users went through the entire process of dashboard design and customization, including data and visual template selection and attribute mapping, to answer multiple questions. We observed this process to verify existing guidelines and collect potentially new ones specifically for designing dashboard systems for novice users.

Q3. What are the differences between novice and expert users? Finally, we wanted to identify differences between the interaction needs and expectations of novices and BI visualization experts. Experts were knowledgeable in understanding which visualization templates and attributes mappings can answer specific questions, but not necessarily in how to interactively construct or customize a dashboard or visualization. In this respect all our users are novices in creation and customization of dashboards, but not in visualization choice and interpretation.

5.1 Participants and Pilot Studies

Overall 15 participants (4 women), ages between 24 and 43 years old (median of 29) took part in our study: 8 were novice visualization and dashboard design users, and 7 were BI dashboard experts. Visualization and BI dashboard novices had no previous knowledge of visualization creation and were only familiar with charts (e.g. pie-charts) seen in everyday life in newspapers or presentations. They reported being exposed to charts on a monthly basis. When asked, only 2 novice users reported that they knew what a dashboard was, and one of them had used the Google analytics dashboard. All novice users were familiar with the concept of data filtering, but at least half did not understand concepts such as drill down-up. Novice users were recruited from a university community, with a wide professional expertise (IT, medicine, and pharmacology). Experts were product users of a BI dashboard provider with different areas of expertise (finance, human resources, analysts, etc). Experts reported using charts and dashboards on a daily basis and were familiar with multiple types of charts, data analysis types (such as correlation, trends, comparisons, etc), and data exploration concepts (filtering, drill-down/up, pivot, roll-up, sorting, etc).

We conducted a pilot study with two novice participants and found that both had trouble understanding the difference between a measure and a category, as well as some data exploration concepts such as data filtering and drill-down/up. We thus decided to include a graphical explanation of these aspects as training in the actual study. We also observed some interface problems: some icons were not clear enough, and were changed to more easily understandable icons and text descriptions.

5.2 Procedure and Apparatus

Participants performed the study in a usability lab using a 15.6" HD screen laptop. Sessions was audio/video recorded and screen captured. One observer was always present to give instructions, explain tasks and observe participants.

Each session lasted between 1.5 to 2 hours. It started with a survey on participant's background experience in computer usage, chart and dashboards, and specific data exploration functionalities such as filtering, drill down/up, etc. As one of our goals is to see how easy it is for novice users to discover functionality and create dashboards with EV, we provided graphical explanations and examples of concepts such as measures and categories, data filtering and drill-down/up, but no information about how to perform them in EV. Next, participants were introduced to a hypothetical situation and dataset (see **Sec 5.3**). We followed a think-aloud protocol, requesting users to vocalize their thoughts and actions while conducting their tasks. This phase lasted between 35-55 minutes. After finishing all the tasks in our scenario, users were asked to complete a Likert type questionnaire to elicit their opinion on specific aspects of EV, and were interviewed to clarify observations made by the experimenter, identify problems users faced, and suggest areas for improvement.

5.3 Tasks and Dataset

Our dataset is an e-fashion dataset (as in [18]), containing 3860 records with 9 measures (e.g. sales revenue, quantity sold) and hierarchical categories (e.g. state/city /store, year/quarter/month, line/category). The dataset was new to both novice and experts so the impact of the dataset and domain on task performance is limited. The dataset is also complex enough to allow complicated analysis questions.

Our tasks were presented within a larger scenario. Participants were told they work in the financial department of a global chain of garment sales, monitoring financial aspects using a dashboard tool. We encouraged participants to think aloud, and if they failed to do so we prompted them to vocalize their insights and expectations.

Based on our pilot, we provided participants with a sheet explaining graphically of the nature of measures and categories, as well as filtering and drill down/up. We also give them a brief explanation of the dataset (what measures and categories exist), and told them to use the dashboard to create and rearrange visualizations.

Five tasks were performed, ordered from simple to complex (**Table 1**). Each task encouraged users to use specific operations and data exploration functionalities such as filtering. Tasks began by loading an existing dashboard and participants were given time to familiarize themselves with the dataset and dashboard.

At the end of the session participants took part in an interview and filled out a questionnaire. The goal of the interview was to clarify observations made during the study, prompt users to identify difficulties/barriers they faced, and suggest improvements. The goal of questionnaire was to rate EV on a 7 point Likert scale on aspects such as its utility, ease of use, overall satisfaction of use, etc.

Table 1. Tasks users performed during each session of the study, and their purpose

#	Task script	Purpose of Task
T1	You are working on sales data for your company with a new dashboard tool (EV). Find and load the dashboard "Sales USA 2004-2006" and explore the new tool.	Create a new EV dashboard and explore its contents.
T2	Your manager calls to request an overview of *Sales Revenue* and *Quantity Sold* in the *Year: 2006* in the *States: California and Texas*	Select visual templates, measures and categories, and data filtering.
T3	He later requests more information on the *States* of *California* and *Texas* in *Year 2006*: 1. Asks for a chart of the top 3 stores generating most *Sales Revenue*. 2. Requests to store a copy of this chart. 3. Asks for a comparison between *Sales Revenue* and *Quantities Sold* for *Items accessories and dresses*. 4. Asks for an alternative visualization of the chart stored in T3.3 to choose from.	Customize charts, using options such as visual template changing, sorting, chart saving and re-use. Conduct different types of common data analysis (e.g. comparisons, extreme values)
T4	Your manager contacts you again to request the *Year* that your company sold most (*Quantities sold*) items from the product *Line: "Leather"*.	Conduct different types of common data analysis (e.g. comparisons, extreme values)
T5	Your manager needs a view showing the U.S. *Sales* for *Years 2004-2006* based on a dashboard image he likes. He asks you to create a new dashboard looking like this one (**Fig 6**) with the data from your dataset.	Design of a complete EV dashboard that incorporates global/local data filters, and conduct layout management

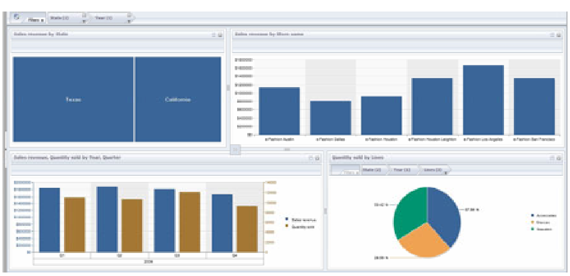


Fig. 6. E-fashion dashboard that participants were asked to replicate

5.4 Data Analysis

We used four different data capturing and analysis methodologies: functionality checklists, observations, questionnaires and open-ended interviews. Observations made during the study were verified, and expanded upon, in a second pass through the recorded material (video, audio and screen captures).

To answer **Q1** on **discoverability** and **usability**, we kept a checklist of functions in the interface and observed how easy it was for participants to discover and use them (results in **Sec 6.1**). Questionnaire answers also contributed towards answering **Q1**.

To answer **Q2** on **guidelines for building dashboard systems** targeting novice users, we noted success events and barriers faced during dashboard creation and customization. These observations are augmented by user comments in the open-ended interview session. Findings are summarized in **Sec 6.2**, and the resulting new guidelines are presented in **Sec 7**. To answer **Q3** on **reactions from novices and experts**, we report explicitly only the differences between the two

6 Findings

We present our main findings on ease of use and discoverability of the EV interface (observed and user self-reporting, **Sec 6.1**), as well as broader findings (based on observations), applicable to dashboard design for novice users in general (**Sec 6.2**). We then provide design guidelines for creating dashboards for novices (**Sec 7**).

6.1 Discoverability and Ease of Use of Interface

All 8 novice (and 7 expert) participants were able to successfully perform all tasks.

To examine if the functionality of EV was easy to use and discover, an experimenter observed the study session and noted difficulties encountered by novice users (8 users). Almost all users easily discovered and performed actions to create new charts (7/8), take snapshots (7/8), manipulate measures, categories and visual templates (7/8), and exploration actions like sorting and drill-down/up (7/8). All novice users applied local filters easily (8/8), although 2 had initially added global filters, as they expected both global and local filters to be added in the same way.

Functionality that novice users had trouble discovering, was chart cloning (4/8), locking (8/8, see section on **Locking** below), and activating chart explore mode (2/8 see section on **Terminology**). Users who did not discover this functionality on their own, were able to complete actions after a small hint from the experimenter.

We examined if users reacted to the interface and functionality differently, by analyzing a 7-point Likert scale questionnaire. Analysis was conducted using a Wilcoxon rank sum test with continuity correction as our samples (Experts, Novices) had different sizes. Overall (**Table 3**), novices ranked EV higher, which can be attributed to the novelty of using a dashboard. Responses were statistically different for: functionality, consistency, and overall satisfaction (all $p<.05$). Experts would have been more satisfied with extra functionality (mentioned next), while both groups had some trouble with representation consistency of global vs. local filters (**Sec 6.2**).

In the follow-up interview both groups of users reported EV was easy to use and customize. They appreciated that customization and data exploration actions became active immediately for quick feedback and verification of their choices. They stated that EV is intuitive and doesn't require training. Finally, mostly our expert users requested additional functionality, like extra customization (e.g. changing colors) and annotation features, and the ability to export dashboards in a printable version.

Table 3. Average ratings of EV for different attributes, by Novice and Expert users (* sig)

	Utility	Functionality	Ease of Use	Consistency	Satisfaction
Novice	5 (std. 1.12)	6* (std. 1.12)	5.5 (std. 1.58)	5.7* (std. 0.83)	6.5* (std. 0.74)
Expert	5 (std. 1.63)	4* (std. 0.75)	5 (std. 0.53)	6.1* (std. 0.95)	5* (std. 0.69)

6.2 General Observations on Dashboards for Novices

6.2.1 Chart Creation and Customization

Chart recommendations. As suggested in previous work on visualization novices, we provided a chart recommendation feature. All participants commented on the value of EV in helping them create appropriate charts fast. Although all users chose charts from the recommendation, only novices experimented with different alternatives, in order to "see what this new visual looks like", "look at a chart I've never seen before". We've had at least 2 users discovering and learning in this way a new visualization that they used often until the end of the study. Nevertheless, in most cases, novice and expert users ended up using familiar visualizations (bar and pie charts) for answering questions ("it looks complex, let's stick with what we have").

Base views and experimentation. We often found that users (novice and experts) created basic charts with measures they thought important (e.g. one for sales over time, one for number of items sold per state, etc). These charts (*base views*), usually 2-3 in number, were used as a starting point for customizing and refining to answer different questions. When customizing a chart, we observed that novices experimented frequently with filters, drill-up/down operations and measures, to "play with things and understand the different options". On the other hand experts attempted to create the appropriate visualization without much exploration.

Chart per question. We observed that both user groups created a single visualization per task question. When asked if they see a benefit in using a dashboard, since one chart was the focus of the analysis at a time, they all mentioned that it "helps keep the context of all data" (benefit identified in BI dashboard literature [9]).

Use of categories. Novice participants often did not recognize the effect of adding hierarchical categories on a chart. In EV when a refined category is added in a chart, it affects aspects such as sorting, aggregation (averages), etc. For example a user had created a chart with sales sorted by year. She then added the category quarter, which is a drill-down category for year. At this point sorting was applied by quarter, but the user still expected sorting to be done by year. Thus automatically calculating metrics based on the finer categories of a hierarchy was not understood by novice users.

Data filtering. As we mentioned EV allows both global dashboard data filtering, as well as local filtering on a single chart. All users found this functionality useful, nevertheless it was often hard for novice users to make correct use of it.

There are 2 creation options for global filters: select part of the data on a chart and then apply the option “filter” which will filter data based on the selected categories and values; or select the option “filter” and then go over the process of selecting on which categories and values to filter. We found that novices generally preferred the second option because they want to make sure they “selected all correct values”. Contrary, experts created global filters with selections that they then refined further.

Another aspect where expert and novice users differed was the presentation of possible filter categories. Experts wanted categories to be presented grouped thematically, e.g. grouping of time (year/quarter), or geographical categories (state/city), while novices preferred an alphabetic presentation.

Users found the existence of both global and local filters useful and relevant when analyzing data as it allows advanced analysis, for example comparison of different products sold by stores for several periods. All participants commented that presenting global and local filters consistently is important, but that local filters should be further highlighted in charts. The EV interface gave such detailed filter views, and users commented on the bulkiness of the filter panels. There is a thus a tradeoff between presenting filter details and space. Finally, we found that sometimes users chose local filters that were in conflict with global ones. This caused problems for several novice users who could not understand why no data was displayed.

6.2.2 Interaction

Linking & coordinated views. Most participants (experts and novices) expect a link between charts. Thus when selecting a portion of one chart to highlight, they expect other charts with the selected data in EV to update accordingly. Nevertheless, the expectation of what “update” means differs across users. Most expert users expected selections on one chart to be reflected as selections in all the others, in the classic brushing and linking approach in coordinated views (default behavior in EV, **Fig.4**). Nevertheless, most novice users expected selections to act as explicit filtering: when partial data was selected in a chart, they expected all charts to update such that only the selected data is reflected throughout the dashboard. This interpretation would result in changes to measures such as averages, sorting, etc for different charts.

Undo as exploration strategy. All participants were familiar with undo/redo, but we found that it was novice participants that made extensive use of undo per chart. Undo was not used as a correction mechanism when mistakes were made, but rather as an exploration aid. Novice users reported that if they wanted to try something and see if their predictions were correct (e.g. drill-down), they could use undo to quickly dismiss their tests (instead of reversing their actions, e.g. drilling-up). They also used undo as a way of returning to *base views* (charts used as starting points for further exploration): from base charts they would drill-down/up or filter briefly to answer a question, and then use undo to return to the original chart.

Clipboard. This feature allows users to create a palette of active charts that they can later reuse. All participants found it useful for storing, but they reported it is most useful for sharing chart snapshots with others. This remark, reinforced by requests for ways to email or export charts to presentations, indicates users also saw the clipboard as a storytelling creation platform for sharing their findings with others. Participants from both groups also requested features such as annotation, putting titles, adding comments, and assigning the current date to a chart. Experts in particular requested stored charts to also clearly and visibly describe context (e.g. dataset, measures, filters, etc). These requests indicate that participants want to use the clipboard to bookmark their exploration process: keep track of interesting charts they explored, comment them, and store the date they conducted their exploration.

Some novice users mentioned they were unsure exactly what is captured when a selection is active on a chart (i.e. does the capture only include the selected data or not). This is an interpretation problem similar to that seen in coordinated views.

Visualization locking. In our prototype we provided functionality for visualization locking (not allowing editing or global filters). Expert participants expected this locking to apply to filters, but also visual properties like size and position in the dashboard. Some novices felt that locking a visualization "creates confusion because it takes out this particular chart from the context of the dashboard". They requested very clear visual representation to indicate locked charts, such as depressed charts.

Terminology. Overall users were satisfied with the terminology and labels used in EV. Some novice users suggested more precise terminology even for some features that are familiar from other UIs. For example our original icon for editing a chart was confusing and novice users were not sure if they were going to edit the title, the chart or the entire dashboard. Thus they often requested tool-tips detailing the function of icons and their context. Finally, while experts were accustomed to English terms (in terms of functionality, dataset measures and categories), this was not true for novices with English as a second language who requested support in their mother tongue.

6.2.3 Design Tradeoffs

The design of EV balances several tradeoffs to support both expert and novice users. Functionality required by experts, such as chart locking, axis flipping, and extended chart recommendations, often confuses. In our work, we attempted to always provide the simpler and most used functionality in a more prominent way. For example options like sorting, filtering, and visualization changing are first in the chart widget drawer, while axis flipping or locking come last. In the chart recommendation list, most common charts always appear first. Moreover, when possible we use reasonable defaults, for example in assigning color and other visual variables to categories. Nevertheless, we often found that functionality required by expert users still confused novices occasionally (e.g. the locking option).

7 Discussion and Design Guidelines

All novice participants in our study were able to successfully perform the required tasks in EV, a system built using guidelines for novice user interaction. Based on our observations we have additional guidelines when designing visualization systems for novice users, especially in the context of dashboards.

Basic training on concepts such as categories: Our novice users were successful in creating visualizations and solving their tasks. But, based on our pilots, they were given small tutorials on the differences between measures, categories, and the nature of basic OLAP and data exploration functionality (drill-down/up, filtering, etc). Thus dashboard interfaces for novices should provide small tutorials of these concepts. This training is particularly important for hierarchical categories. Novice users need to know if metrics (averages, sorted lists, etc) are recalculated in charts when adding finer hierarchical categories. Based on user comments we recommend either not re-calculating metrics unless users request it, or recalculating per category on demand.

Considerations in coordinated views: When creating a dashboard, novice users assume that charts are linked by default, and expect brushing and linking functionality. Its implementation for single data sources where charts share measures and categories is fairly easy. Nevertheless, if measures and categories need to be matched across data sources, then matching approaches need to be provided [21,32]. When providing coordinated views, designers should remember that novice users confuse linked selections of visual queries to applying global data filters based on the selections. To help disambiguation, they need clear visual differentiation between selection for linking and for data filtering. A linking selection should highlight selections without affecting the visual representation of unselected data; a filtering selection should also alter the visual representation of the unselected data (e.g. fade them out) to indicate that all data are being altered due to the selection.

Extensive exploration aids: Novice users created a small number of *base view* charts, most often of familiar visualization templates, that they later copied and refined to answer questions. When possible, such basic views should be provided. Novices also made extensive use of chart undo/redo as an exploration aid, thus this functionality should be provided on a per chart basis. Furthermore, clipboards were seen as good ways to share data for storytelling and communication, as well as bookmarks of the exploration process. For clipboard snapshots to be useful this way, they require functionality such as annotations and labeling, as well as ways to report their metadata, such as creation date and details of the context of the chart (i.e. dataset, active filters, etc). Finally, although locking charts can be useful during long analysis (as indicated by experts), this locking should be very clearly visible to not confuse novices (e.g. fading the entire locked chart, or making it appear as depressed).

Data filtering considerations: Data filters, although very useful for data exploration, can be challenging for novice users. We found that novices preferred to create global filters from scratch, select categories, values, etc. While experts preferred to create them graphically (through visual selections) and then refine them. Both options should be available. When selecting categories to filter, we recommend presenting them in thematic groups (expert choice) that promotes dataset learning. But this organization should be supplemented by text search functionality. Novice participants want filter details to always be visible (both local and global), while experts found they take up too much space. Designers should consider an auto-hide, or on/off visual representation of filters. Approaches such as interactive legends [17] could also reduce clutter of local filters, since most charts include labels of some sort. For both groups, being able to distinguish between global and local chart filters is important, and it should always be clear which categories and values contribute to existing filters. Finally, since novice users have trouble predicting conflicts in global and local filters, it is important to clearly mark such conflicts.

Language & terminology: Novice users are often not aware of international terms for different charts, metrics or functions. Thus local language support is important.

Moreover, buttons, labels and icons (even universally understood ones) are not enough to provide novices with adequate information for their behavior in the context of dashboards. Detailed explanations need to be provided on demand (e.g. tool-tips).

8 Conclusions

We have presented Exploration Views, a prototype system that allows users to easily create, customize and interact with Business Intelligence visualization dashboards. EV was build following previous guidelines for novice visualization users. Through a user-study with novice and expert visualization users we verified that EV enables novices to create dashboards for answering complex questions. We found some differences in how experts and novice users interact with the interface, reported in our study. More importantly, based on our observations we provide new guidelines that augment previous work on designing for visualization novices, especially in the context of interactive visualization systems in the form of dashboards.

Of course further work is needed to gain more insights in how novices use dashboards. As study participants were not data experts, it could be interesting to see if our findings extend to data experts. Moreover, it would be interesting to observe novice user exploration behavior in a long-term study. Finally, our work focused more on the user interface design and use in a visualization environment for novices, and less on the insights gained from the data. Clearly this an avenue of future work.

Acknowledgments. We'd like to thank our users for their time and comments, Alexis Naibo and Didier Bolf for their support, and our reviewers for their useful suggestions. This research was funded by SAP BusinessObjects Innovation Center and SAP Research.

References

1. Card, S.K., Mackinlay, J.D., Shneiderman, B.: Readings in information visualization: Using vision to think. Morgan Kaufmann Publishers, San Francisco (1999)
2. Capra, R., Marchionini, G.: The relation browser tool for faceted exploratory search. In: Proc. JCDL 2008, pp. 420–420. ACM Press, New York (2008)
3. Chiang, A.: Creating Dashboards: The Players and Collaboration You Need for a Successful Project. Business Intelligence Journal 14(1) (2009)
4. Choosel (2009), http://code.google.com/p/choosel/
5. Cubrarnic, D.: Polstar: Assisted navigation for exploring multi-dimensional information spaces. In: Huma-Computer Information Retrieval, HCIR (2008)
6. Cutroni, J.: Google Analytics. O'Reilly Media, Inc., Sebastopol (2007)
7. CVOM patent, http://www.freshpatents.com/Apparatus-and-method-for-data-charting-with-an-extensible-visualization-library-dt20080731ptan20080180458.php
8. Dundas dashboard, http://www.dundas.com/Microsite/Dashboard
9. Few, S.: Show Me the Numbers. Analytics Press (2004)
10. Few, S.: Information Dashboard design. O'Reilly, Sebastopol (2006)
11. Foody, P.: User-Centered Business Intelligence. Business Intelligence Journal 14(4) (2009)
12. Grammel, L., Tory, M., Storey, M.-A.: How Information Visualization Novices Construct Visualizations. IEEE InfoVis - Transactions on Visualization and Computer Graphics 17(6), 943–952 (2010)

13. Hearst, M.: User Interfaces and Visualization. In: Baeza-Yates, R., Ribeiro-Neto, B. (eds.) Modern Inf. Retrieval, ch. 10 (1999)
14. Hearst, M.: UIs for Faceted Navigation, Recent Advances and Remaining Open Problems. In: HCIR 2008 (2008)
15. Heer, J., van Ham, F., Carpendale, S., Weaver, C., Isenberg, P.: Creation and collaboration: Engaging new audiences for information visualization. In: Kerren, A., Stasko, J.T., Fekete, J.-D., North, C. (eds.) Information Visualization. LNCS, vol. 4950, pp. 92–133. Springer, Heidelberg (2008)
16. Heer, J., Viégas, F.B., Wattenberg, M.: Voyagers and voyeurs: Supporting asynchronous collaborative information visualization. In: Proc. of the Conference on Human Factors in Computing Systems (CHI), pp. 1029–1038
17. Henry Riche, N., Lee, B., Plaisant, C.: Understanding Interactive Legends: a Comparative Study with Standard Widgets. In: Proc. Eurovis 2010, vol. 29(3), pp. 1193–1202 (2010)
18. Mahyar, N., Sarvghad, A., Tory, M.: Roles of Notes in Co-located Collaborative Visualization. In: Workshop on Collaborative Visualization on Interactive Surfaces (CoVis 2009) (October 2009)
19. Marchionini, G.: Toward Human-Computer Information Retrieval. Bulletin of the American Society for Inf. Science (2006)
20. Nielsen, J.: Usability Engineering, pp. 115–148. Academic Press, San Diego (1994)
21. North, C., Shneiderman, B.: Snap-Together Visualization: A User Interface for Coordinating Visualizations via Relational Schemata. In: Proc. AVI 2000 (2000)
22. OLAP and OLAP Server Definitions. The OLAP Council (1995)
23. Oracle BI 10g dashboard, `http://www.oracle.com/technology/products/bi/enterprise-edition-platform-components.html`
24. Raskino, M., Lopez, J.: Early Findings From the 2010 Gartner CEO and Business Executive Survey (December 2009)
25. Roberts, J.C.: State of the Art: Coordinated & Multiple Views in Exploratory Visualization. In: Proc. of the Fifth international Conference on Coordinated and Multiple Views in Exploratory Visualization, pp. 61–71. IEEE Comp. Society, Los Alamitos (2007)
26. SAP Business Objects Xcelsius, `http://www.sap.com/solutions/sapbusinessobjects/large/business-intelligence/dashboard-visualization/xcelsius-enterprise/index.epx`
27. Sorenson, M.: Dictionary definition of dashboards, SearchCIO.com, February 25 (2002)
28. Spotfire, `http://spotfire.tibco.com/`
29. Tableau, `http://www.tableausoftware.com/products/desktop`
30. Willhalm, T., Popovici, N., Boshmaf, Y., Plattner, H., Zeier, A., Schaffner, J.: SIMD-Scan: Ultra Fast in-Memory Table Scan using on-Chip Vector Processing Units. In: VLDB 2009 (2009)
31. Viegas, F.B., Wattenberg, M., van Ham, F., Kriss, J., McKeon, M.: ManyEyes: a Site for Visualization at Internet Scale. IEEE Transactions on Visualization and Computer Graphics 13(6) (November 2007)
32. Weaver, C.: Building highly-coordinated visualizations in improvise. In: IEEE Symposium on Information Visualization (InfoVis 2004), pp. 159–166 (2004)
33. Williamson, C., Shneiderman, B.: The Dynamic HomeFinder: Evaluating Dynamic Queries in a Real-Estate Information Exploration System. In: Proceedings of the 15th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 339–346. ACM, New York (1992)
34. Woods, D.: Visual momentum: a concept to improve the cognitive coupling of person and computer. Int. J. Man-Mach. Stud. 21(3), 229–244 (1984)
35. Zelazny, G.: The Say It With Charts Complete Toolkit. McGraw-Hill, New York (2006)

Patient Development at a Glance: An Evaluation of a Medical Data Visualization

Margit Pohl[1], Sylvia Wiltner[1], Alexander Rind[2,3], Wolfgang Aigner[2,3], Silvia Miksch[2,3], Thomas Turic[3], and Felix Drexler[4]

[1] Research Group Human Computer Interaction,
Vienna University of Technology
margit@igw.tuwien.ac.at, s.wiltner@aon.at
[2] Institute of Software Technology & Interactive Systems,
Vienna University of Technology
{rind,aigner,miksch}@ifs.tuwien.ac.at
[3] Department of Information and Knowledge Engineering (ike),
Danube University Krems
[4] Landesklinikum Krems, Austria

Abstract. This paper describes the results of an evaluation study of a prototype for the visualization of time-oriented medical data. Subjects were nine physicians. The prototype combines well-known visual representation techniques and extensive interaction techniques. The aim of the study was to assess the system's usability and whether the prototype solved relevant problems of physicians in hospitals. It was found that one of the great advantages of the system was that it allowed physicians to see the development of the patients at one glance. It was also shown that users clearly preferred an easy to learn and understand design and familiar visualizations.

Keywords: interaction techniques, user study, time-orientated data, visual exploration, medical data.

1 Introduction

Physicians treating patients with chronic diseases usually have to deal with large amounts of time-oriented data. It is often difficult to get an overview of this data. Information Visualization (InfoVis) might make it easier to derive relevant insights from this data and get a quick idea of what factors influence the patients' development over time. During the project described in this paper, a prototype of an InfoVis system to solve this problem was developed. This prototype is highly interactive to allow the physicians to accommodate it to their needs and to provide them with different views on the data. It was evaluated with nine physicians. The goal of this evaluation was to find out whether the prototype solved the needs of the physicians and whether the visualizations and interactive features of the systems were easy to learn and useful. The results of this work are not only relevant for the system at hand but also for similar systems that represent large amounts of time-oriented medical data using interactive features and a linear time axis as a metaphor.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 292–299, 2011.

2 Related Work

Computers play an increasing role in health care, e.g. in electronic medical records. Sainfort et al. [12] report that possible advantages of electronic medical records include better access and availability of patient information for physicians and the availability of full-text search. Patel et al. [6] point out that electronic medical record systems might lead to a lack of time-oriented information concerning the development of patients and that novel features of such systems might be used to overcome this problem. Patel et al. [5] describe thinking and reasoning in medical diagnosis in great detail. They argue that these processes are influenced by the specific affordances of computer-based systems. The design of such systems is, therefore, of great importance.

Making medical records visually accessible is a prominent research area. Therefore, we can only present few, most relevant contributions here and refer the readers to a more extensive review [1].

The Graphical Summary of Patient Status (GSPS) [8] and LifeLines [7] are seminal works in this area. While GSPS and Clinical Care Patient Data Visualization (CPDV) [4] show numerical variables in scatter plots, LifeLines displays nominal variables in a similar fashion to our timeline chart. Midgaard [2] extends LifeLines with a semantic zoom chart technique that shows more details as the chart is enlarged. These systems offer varying degrees of flexibility. Most of them, with the exception of Midgaard, rely on a single visualization technique for either numerical or nominal variables. Further, most of them, with the exception of CPDV, do not allow the physician to rearrange, add, or remove variables from the display.

3 Analysis of Medical Requirements

For a thorough analysis of users, tasks, and data we involved physicians in a user-centered design process from the very beginning, as described in [11]. We developed our prototype based on the following requirements, which we elicited from introductory interviews of five physicians and contextual observations: (1) Simple user interface and visualizations, which are particularly clear and allow for unambiguous insights; (2) flexibility to work with various medical variables and combine different groups of variables; (3) time-oriented data representing patient development over time; (4) multiple patients; and (5) interactivity through a variety of interaction techniques [11]. Furthermore, we had repeated workshops during the development phase to discuss and refine the design based on feedback of InfoVis experts, usability experts, and physicians.

In order to provide a realistic medical scenario for the design process and the user study, we worked with data and tasks from a diabetes outpatient clinic. The data is composed of medical test results and therapy prescriptions that are collected at check-up examinations scheduled about every two months. Though, the examination intervals and the provided data items may vary largely depending on the patient's condition. Physicians are interested in whether the condition of a diabetes patient improves or worsens. How the improvement of some variables relates to the development of others is of particular importance. They also need to know which types of therapy a patient already had and how these affected medical tests.

4 Design of the VisuExplore System

Next, we describe how we combined simple and well-known visual representation techniques with extensive interaction techniques that meet four of the five medical requirements presented above. Visually exploring the developments and relationships of multiple patients imposes a different set of tasks to support. Therefore, a separate system for multiple patients has been developed and is reported in [10].

VisuExplore (Fig. 1) displays medical variables in multiple diagrams, which are aligned along a horizontal time axis. Thus, it encodes time by positioning on a common scale, which is perceptually effective [3] and should be easy to learn. The visual representation techniques used in the diagrams are well known and quite simple. To allow for flexibility VisuExplore provides several representation techniques: Line plots and bar charts for numerical variables and event charts and timeline charts for nominal variables. A diagram may also display two or more variables. Furthermore, we extended the system with five advanced representation techniques: a semantic zoom chart (cp. [2]), a step chart, a silhouette graph, a horizon graph (cp. [9]), and a document browser. Though, in the user study (Sect. 5) we only evaluated simple techniques because the introductory requirements analysis called explicitly for simple visualizations.

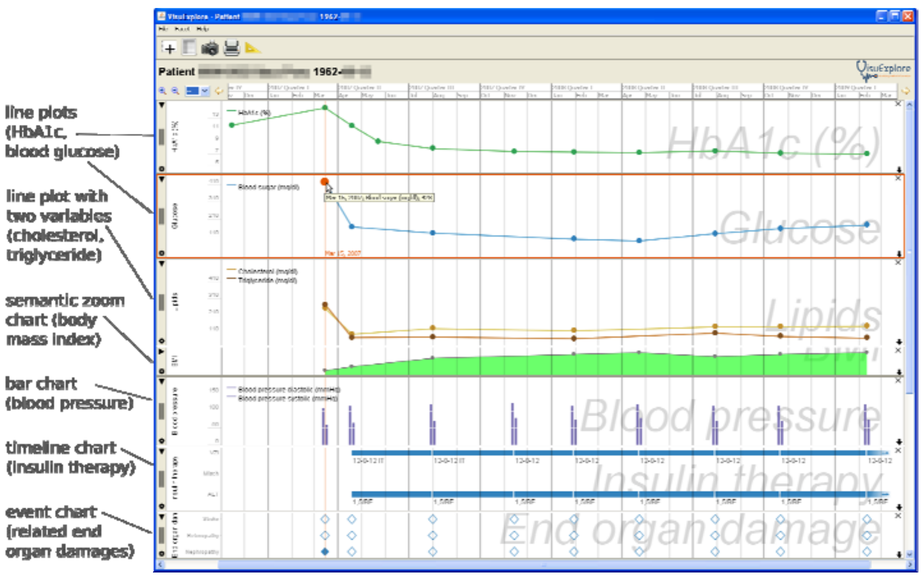


Fig. 1. Exploring a medical record using VisuExplore

VisuExplore starts with a predefined set of diagrams, which is customizable to accommodate for different medical specialties. In our diabetes scenario we provided a line plot of blood glucose, a second line plot of cholesterol and triglyceride, a timeline chart of insulin therapy, etc. But this set is just an initial aid for physicians, as VisuExplore offers a number of interaction techniques to adapt the diagrams.

Physicians can create new diagrams and close existing ones. They can drag diagrams to change their order, and thus compare developments in two diagrams more easily. They can toggle each diagram between a large and a collapsed view or drag its border to change the height continuously. Thereby our representation techniques adapt to their available screen space and show more or less detail. Each diagram has a control panel for its settings. For example, a line plot setting specifies whether the y-axis zooms to the values of the current patient or extends to a domain-specific value range.

If users are interested in a specific time interval, they can zoom in continuously with the mouse or select from a predefined set of meaningful zoom steps such as "1 year" or "3 months". They can scroll horizontally across time as well as vertically along diagrams.

Further interaction techniques of VisuExplore support physicians in the exploration of the medical data. As they move the mouse across the window, a vertical line follows the cursor to help them find co-occurring items. Using tool tips and a table panel they can read exact data values. The table panel shows all items of a diagram textually and can be sorted by any column (e.g. variable value). Physicians can select one or more data items either in the diagram or in the table panel, whereupon selections are linked. Finally, VisuExplore provides a tool for physicians to measure the duration of a time interval or the time difference between two items, especially to judge a possible cause–effect relationship. With the measure tool, they can drag a metaphorical tape measure from a reference point to the current mouse location, which works as well across multiple diagrams (Fig. 2). Similar tools are often found in geospatial visualizations, but are uncommon for time-oriented data.

5 Description of the Investigation

The aim of the user study was, on the one hand, to assess the usability of the system and, on the other hand, to check whether the system aids in solving relevant tasks of medical doctors in hospitals. The most relevant research questions were:

1. Which of the physicians' tasks can be supported by the proposed system?
2. In which departments of the hospital should the system be used?
3. Which forms of interaction are most useful?
4. Which of the proposed visualizations for the variables are appropriate?

The subjects were nine physicians from two hospitals. Seven of them were women, two men. This reflects the fact that many of the physicians in the two hospitals taking part in this study are women. Most of them were under 40 years of age, only one was between 50 and 60. Four of them were specialists in diabetes care, the others were familiar with diabetes. We selected experts in fields where the tool might realistically be applied, especially in fields with large amounts of time-dependent data (e.g. in the treatment of pulmonary diseases). The subjects were not only early adopters but had various degrees of previous knowledge of computer usage.

It is well known that it is difficult to do user testing with professionals, especially because of their time constraints. Therefore, we tried to organize the investigation in such a way to get the most out of the limited time the physicians were available. First, we gave them a short introduction into the system (10 min), then they had to solve three tasks. An example for such a task is: "Please look at the data of three patients (A, B, C), especially at blood sugar, cholesterol and body mass index. Can you see any relationships between the three parameters for these three patients?" These tasks were developed together with a diabetes expert from the hospital. The time subjects spent working on the tasks was approximately 20 minutes. In the end, we conducted a short qualitative interview with open-ended questions (planned were 10 minutes, but some interviews lasted for 30 minutes). The interview partners sometimes indicated relevant features on the screen and gave us concrete examples to substantiate their views.

The main method used in this study to answer the research questions was interviews. The results presented in this paper are based on these interviews. The questions asked were basically refined versions of the research questions outlined above. We asked them specifically to comment on the advantages and disadvantages of all available interaction methods. The interviews were then transcribed. The interpretation of such interviews is often difficult because interview partners utter conflicting statements (e.g. they do not appreciate a certain feature, but would use it if improved in well-defined ways). Therefore, quantitative scores often do not reflect such results. In the study, only the simple representation and interaction techniques were tested.

6 Results

In this section, we give an overview of the results of the interviews with the physicians.

6.1 Advantages of the System (Research Question 1)

A great advantage of the system mentioned by all participants in the study is the fact that users can see the development of the patients at one glance. One very important input mentioned by the physicians in the requirements analysis was that for the treatment of patients with chronic diseases they have to read many reports on diagnostic findings. This may take up much of their time, and it is a great challenge to derive an overall impression of the patients' development from these reports. The VisuExplore system can support physicians considerably in this activity. It apparently solves the physicians' problem with information overload.

In addition, most of the participants mentioned that the system was easy to learn and understand. The visualization methods were found to be intuitive and clear.

6.2 Potential Areas of Application (Research Question 2)

Most of the physicians agreed that the VisuExplore system could best be applied for outpatient care of chronic diseases (diabetes, treatment of cardiac insufficiency, cancer, etc.). According to the interviewed physicians, it would not be so useful for inpatient treatment because most of these patients stay in contact with the hospital for much shorter periods of time. Another possible application mentioned by some subjects is that the system could be used to communicate to the patients the state of their health. Some physicians thought that the patients might be more impressed by a visual representation of the data than by just showing bare numbers. Furthermore, it was mentioned that it might also be used for teaching purposes.

6.3 Interaction (Research Question 3)

Resizing Diagrams (Fig. 2): In the VisuExplore system, it is possible to increase/decrease the height of the diagrams. The physicians' attitude to this possibility was mixed. Three found this quite useful, others thought this was not necessary or even misleading (four noted that when decreasing the height of the diagrams, differences between values are also considerably decreased giving a wrong impression of the patients' development). In contrast, some physicians liked this possibility because in they could get a better overview of all necessary values.

Moving Diagrams: Seven of the physicians explicitly liked this possibility. They commented that they found it interesting to create their own view of the data so that they could use the system most efficiently. Some of them also mentioned that they would create this view once and then stick to it for most cases, but that this view would be different for every physician.

Measure Tool (Fig. 2)*:* Concerning the measure tool, the physicians' reactions were mixed. Only three really liked this feature and others found this unnecessary or confusing. One subject proposed that the tool should calculate the mean value of the data in the selected time interval.

Pan and Zoom: Both were found to be interesting under certain conditions, especially when there were many data items with changing values. However, it must be mentioned that in many cases the line plots in combination with a table showing the concrete values can provide enough information to get a clear overview.

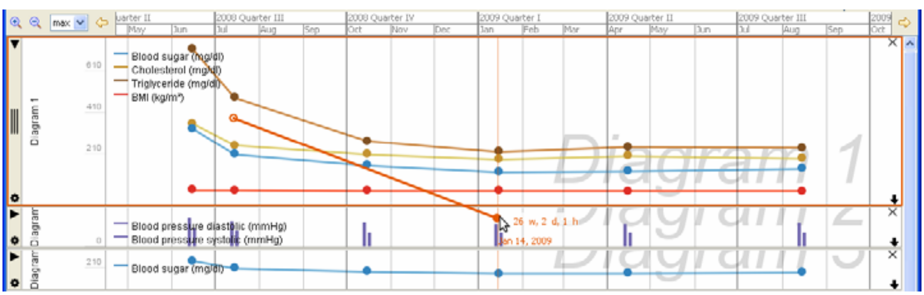


Fig. 2. The first diagram combines several variables. The second and third diagram are resized. The measure tool is shown (slanted orange line) going from the first to the second diagram.

Scrolling: We observed that the activity physicians adopted most was scrolling. This is partly due to the visualization itself (one diagram below another), but we also observed that they seemed to prefer scrolling to closing or resizing diagrams.

6.4 Visualizations (Research Question 4)

Several Variables in one Diagram (Fig. 2)*:* There were also mixed reactions concerning this issue. Four physicians preferred to combine several variables represented by line plots (e.g. cholesterol, triglyceride, BMI, and blood sugar) in one diagram because of the compact form and the possibility to see everything at a glance. Other physicians found this confusing, especially because of the different scales involved. Furthermore, BMI has a different unit of measurement. In addition, they found too many lines in one diagram confusing. There were, apparently, different approaches to solve the given tasks. It should be pointed out, however, that all physicians found plausible solutions for the tasks while using different methods for reaching these solutions.

Methods of Visualization (Fig. 1)*:* The system offered several methods of visualization (line plots, bar charts, etc.). There was a general consensus among physicians that line plots were the best form of representation. They argued that line plots are well known and can therefore be interpreted easily. Line plots can show the development in a convincing manner, outliers can be identified easily. Bar charts for blood pressure are fine, but even here they would have preferred line plots.

In general, well-known visualizations were preferred. Because of their time-constraints and workload, physicians are very reluctant to familiarize themselves with unknown forms of visualization.

7 Conclusion and Future Work

Our study indicates that the system is considered to be useful for supporting the therapy of patients with chronic diseases like diabetes.

During the requirement analysis physicians expressed their wish for a system that is easy to understand and learn. It was often mentioned during the user study that our system satisfies these needs and that it is very simple and intuitive.

Physicians also showed a clear preference for line plots because they were familiar with this visualization and could therefore easily and quickly interpret them. Other forms of visualizations should probably be used sparingly.

There are other features of the system where flexibility seems to be welcome and strong personal preferences of the physicians could be observed (e.g. combining several variables in one line plot, moving and resizing diagrams). It was also stated that personalization of the interface according to individual needs is an issue.

The evaluation study indicates that four of the five requirements described in Sect. 3 were fulfilled by the system. The interface is simple and easy to learn, the system is flexible and can be adapted to the users' needs, time-oriented data can be understood at a glance, and a reasonably level of interactivity is supported. The fifth requirement (comparing multiple patients) is fulfilled by another visualization system called TimeRider, which can be combined with VisuExplore [10].

The investigation described in this paper is part of a larger, ongoing study. The aim of that study is to clarify in more detail what strategies users of information visualizations adopt and to identify interaction patterns so that information visualizations can be designed to support users more efficiently.

Acknowledgments. This work was supported by the Bridge program of the Austrian Research Promotion Agency (project no. 814316) and conducted in cooperation with Danube University Krems, Vienna University of Technology, NÖ Landeskliniken-Holding, Landesklinikum Krems, NÖGUS, systema Human Information Systems.

References

1. Aigner, W., Kaiser, K., Miksch, S.: Visualization techniques to support authoring, execution, and maintenance of clinical guidelines. In: ten Teije, A., Lucas, P., Miksch, S. (eds.) Computer-based Medical Guidelines and Protocols: A Primer and Current Trends, pp. 140–159. IOS, Amsterdam (2008)
2. Bade, R., Schlechtweg, S., Miksch, S.: Connecting time-oriented data and information to a coherent interactive visualization. In: Proc. ACM SIGCHI Conf. Human Factors in Computing Systems (CHI), pp. 105–112 (2004)
3. Cleveland, W.S., McGill, R.: Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. J. Am. Statistical Assoc. 79(387), 531–554 (1984)
4. Faiola, A., Hillier, S.: Multivariate relational visualization of complex clinical datasets in a critical care setting: A data visualization interactive prototype. In: Proc. Int. Conf. Information Visualisation (IV), pp. 460–468. IEEE, Los Alamitos (2006)
5. Patel, V.L., Arocha, J.F., Zhang, J.: Thinking and Reasoning in Medicine. In: Holyoak, K.J., Morrison, R.G. (eds.) The Cambridge Handbook of Thinking and Reasoning, pp. 727–750. Cambridge University Press, Cambridge (2007)
6. Patel, V.L., Kushniruk, A.W., Yang, S., Yale, J.-F.: Impact of a Computer-based Patient Record System on Data Collection, Knowledge Organization, and Reasoning. J. Am. Med. Inform. Assoc. 7(6), 569–585 (2000)
7. Plaisant, C., Mushlin, R., Snyder, A., Li, J., Heller, D., Shneiderman, B.: LifeLines: using visualization to enhance navigation and analysis of patient records. In: Proc. AMIA Symp., pp. 76–80 (1998)
8. Powsner, S.M., Tufte, E.R.: Graphical summary of patient status. Lancet 344(8919), 386–389 (1994)
9. Reijner, H.: The development of the horizon graph. In: Proc. Vis. 2008 Workshop From Theory to Practice: Design (2008)
10. Rind, A., Aigner, W., Miksch, S., Wiltner, S., Pohl, M., Drexler, F., Neubauer, B., Suchy, N.: Visually exploring multivariate trends in patient cohorts using animated scatter plots. In: Proc. Int. Conf. Human-Computer Interaction, HCI-I (forthcoming, 2011)
11. Rind, A., Miksch, S., Aigner, W., Turic, T., Pohl, M.: VisuExplore: gaining new medical insights from visual exploration. In: Hayes, G.R., Tan, D.S. (eds.) Proc. Int. Workshop on Interactive Systems in Healthcare (WISH@CHI2010), pp. 149–152. SIGCHI (2010)
12. Sainfort, F., Jacko, J.A., Booske, B.C.: Human-Computer Interaction in Health Care. In: Jacko, J.A., Sears, A. (eds.) The Human-Computer Interaction Handbook, pp. 808–822. Lawrence Erlbaum Associates, Mahwah (2003)

Evaluation of HaloDot: Visualization of Relevance of Off-Screen Objects with over Cluttering Prevention on Mobile Devices

Tiago Gonçalves, Ana Paula Afonso,
Maria Beatriz Carmo, and Paulo Pombinho

Faculdade de Ciências da Universidade de Lisboa, Portugal
{tgoncalves,ppombinho}@lasige.di.fc.ul.pt,
{apa,bc}@di.fc.ul.pt

Abstract. The complexity of presenting and exploring large amounts of graphical data, on mobile devices, increases due to their small screen size. To mitigate this problem several approaches have been proposed to give clues about objects that are located off-screen. In this paper we present a user study comparing the Halo off-screen visualization technique with HaloDot, our approach that aims to improve direction awareness, as well as, relevance of off-screen objects, and to avoid cluttering of Halos. The study shows that searching and pointing relevant Points of Interest (PoI) can be achieved faster than with Halo and that the proposed aggregation method is useful.

Keywords: Visualization, Mobile Devices, Off-Screen Objects, Relevance.

1 Introduction

Presenting and exploring large amounts of graphical data on small screens are key research topics. To search Points of Interest (PoI) in large maps on mobile devices, panning and zooming can be used to explore surrounding areas that are not visible on screen. However, as these techniques are cognitively complex and frequently disorient the user, different approaches to give visual clues of off-screen objects have already been proposed [1, 2, 3, 4, 5].

Another important issue in mobile visualization is the development of mechanisms to show the most relevant information to the user, reducing the amount of information displayed to avoid cluttering. The relevance in a mobile context should capture not only the location of an object, but also other contextual factors, such as, temporal constraints and properties or attributes of an object. Research approaches in this topic aim to establish an appropriate distance function that integrate several contextual factors beyond the location, namely semantic and temporal relevance [6,7].

In this paper we describe and evaluate our approach, HaloDot that enriches the Halo off-screen visualization technique [1], aiming to improve direction awareness of off-screen objects, to give hints about their relevance, and, to prevent cluttering when there are a large number of off-screen objects.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 300–308, 2011.

This paper is organized as follows. Section 2 describes related work about off-screen visualization techniques and relevance visualization. Sections 3 present the HaloDot technique. Section 4 describes the experimental evaluation and reports the results. Finally, section 5 points out our conclusions and future work.

2 Related Work

There are several approaches to provide hints to the existence and location of objects that are not visible on the display area. They consist of graphical representations, such as, lines, arcs or arrows, disposed along the borders of the display area to convey information about the distance and direction of off-screen objects. Burigat et al. call them Contextual Cues [8]. Some examples are arrows, Scaled and Stretched arrows [2], City Lights [3], Halo [1], EdgeRadar [4] or Wedge [5]. Halo is one of the most known techniques. It consists in surrounding the off-screen objects with rings, which are just large enough to reach into the border region of the visible area. Based on the visible portion of the ring, users can infer the location and the direction of the object at the center of the ring, taking into account the arc position and arc curvature [1] (Fig.1 left).

The Halo has been compared with other off-screen visualization techniques and is considered as a successful technique for awareness of the presence of off-screen objects [2, 4, 5]. However, these studies have also observed that Halo had lower accuracy in some tasks caused either by the underestimating of the distances or a Halo cluttering problem.

In the initial phase of our work [9], we observed that none of the visualization techniques of off-screen objects conveys their relevance. The concept of relevance and how to represent it has also been subject of various research studies. It is not enough to select the most relevant objects but it is essential that the filtered objects show their relevance values [10]. Reichenbacher has suggested the use of “warm” colors, like red and orange, to represent more relevant spots, while the less relevant ones would be represented with “colder” colors. Such practice is also mentioned in [11], that states that colors can be used to represent various meanings, one of them temperature (warm = red, cold = blue). Wolfe J.M [12] also points out the importance of color along with other attributes, such as, motion, orientation and size, to guide users’ attention.

We aim to provide visual clues (based on color and transparency attributes) to convey information about the relevance and the distance of off-screen objects, i.e., the distance between its location and the area visible on-screen. Moreover, we want to avoid cluttering. We started our work using the Halo technique [4] and a function that returns a value of the relevance of a PoI belonging to [0, 1]. The value of the relevance of each PoI, is calculated according with the user preferences and his geographic position [7]. This means that the relevance takes into account the distance between the geographic location of the user and the PoI, which is different from the distance represented by the Halo’s arc.

3 HaloDot Technique

We have designed HaloDot that adapts and enhance the Halo interaction technique to satisfy the relevance and direction awareness and prevention of the clutter problem when a large number of off-screen objects are presented. In the next sections we describe our solutions and improvements applied to the Halo technique.

3.1 HaloDot: Improving the Awareness of Direction

Halo provides location and the distance of the off-screen object based on the size and shape of the visible portion of the arc. Since the distance is already provided by the arc's size, we decided to enrich the Halo indirect direction representation given by the arc's position. We have drawn the arc of the Halo with a small circle at the point of intersection between the Halo's arc and the intrusion border, i.e., the inner limit of the area where the Halos are visible. This approach combines Halo with the direction provided in City Lights technique. Fig. 1 (right) shows an example of this small adaptation, called HaloDot.

3.2 HaloDot + Color + Transparency: Awareness of Relevance

Let us consider the following scenario: a user A standing in the center of the visible area wants to find restaurants with preference to Italian ones. Suppose that the result is the off-screen Italian restaurant 1 and the Japanese restaurant 2. Although, object 1 and 2 are equidistant from the user A, restaurant 1 is more relevant than 2, because the user prefers Italian restaurants. As shown in Fig. 1 left, the Halo technique does not give awareness of the relevance of off-screen objects.

We have decided to use color to represent the relevance of the various off-screen objects, since color is known as a good attribute to guide people's attention. Using a "warm-cold" analogy [10,11], we have decided to color the most "relevant HaloDots" with red (warm/hot), the less relevant with blue (cold) and the objects with an intermediate relevance with purple (tepid).

Fig.1 (right) shows the use of this approach applied to the above scenario. The most relevant object is represented with the red HaloDot and the less relevant is represented with a blue and more transparent HaloDot.

Fig. 1. An example of HaloDot with color. Original Halo (left) and HaloDot (right).

However, our approach to represent the relevance brings another problem related with the color visibility of the most relevant object. Suppose that the most relevant object 1 (Italian) is further away than the object 2 (Japanese). The original Halo uses transparency to deal with distance [1] and the arc's transparency grows according with the distance of the user to the off-screen object, which means that the arc of the most relevant object will be less visible than the arc of the least relevant (Fig. 1 left).

We decided to apply a minimum transparency level, so even objects that are too far away still have a visible HaloDot. Assuming that the more visible, the more relevant an object is [10], if the transparency level was selected only based on the distance of the off-screen object to the visible area, there would be the risk that, if a relevant (red) object off-screen was further away than a less relevant (blue) one, the second HaloDot's arc would be more visible. This could induce the user to pick the wrong object. To avoid this, the transparency level is also dependent on the object's relevance. An interval of minimum and maximum transparency is set according with the object's relevance. This way, a relevant object will always have a more visible HaloDot than a less relevant one (Fig. 2 right).

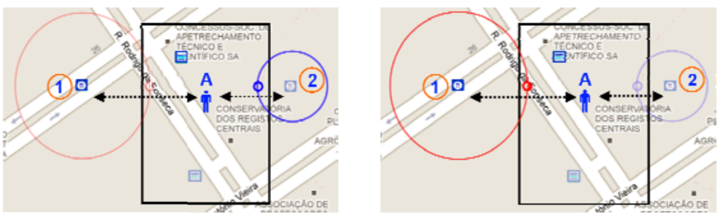


Fig. 2. HaloDot without (left) and with (right) color+transparency to show relevance

3.3 HaloDot Aggregation + Numbering: Cluttering Prevention

As mentioned by several authors [1, 2, 5] one of the Halo's limitations is the cluttering problem when a large number of off-screen objects are presented. Although the use of the color and transparency characteristics of HaloDot could reduce clutter by allowing users to visually separate the most relevant, when the number of off-screen objects is very large the visualization could be very difficult (Fig. 3 left).

In resemblance of what is done with on-screen icons cluttering [7], we propose an aggregation approach to mitigate the clutter problem when arcs density is high (Fig. 3 right). We consider the existence of a hypothetical grid, based on geographic coordinates, overlaying the map, which divides the geographical space into cells. The grid has a default size, though it can be changed by the user. Unlike the previous situation, where a HaloDot would be drawn for each off-screen object, a HaloDot will be drawn for each cell with at least one object. This means that, in the worst case, we will have as many HaloDots as cells. The color and transparency shown by a HaloDot corresponds to the most relevant object it represents.

Fig. 3. Halo (left) and HaloDot with aggregation and number clues (right)

To show the number of objects represented by a HaloDot with aggregation we have investigated several approaches. We tried to change the thickness of the arc and/or the point of intersection based on the number of objects represented. However, these demonstrated being very intrusive, even incomprehensible. We have developed another solution that gives textual information, the number of off-screen objects it represents, near the point of intersection with the intrusion border (Fig. 3 right).

To overcome the cluttering problem at the corners, in analogy with EdgeRadar, where the corners represent a larger off-screen area than the borders [4], we have decided to merge all HaloDots on the corners. Although this means that the HaloDots at the corners might represent more points, since they correspond to bigger cell areas and that the aggregations may change by panning the map, we believe that this will improve the technique, since it greatly reduces the overlap and the intrusiveness of the HaloDot, therefore, improving interaction.

Even with this merging, there is the risk that some HaloDots, located in different cells, have centers with a close latitude or longitude, meaning they can overlap the HaloDots and their textual information. To solve this problem we have considered two approaches: to aggregate all HaloDots that are in cells arranged orthogonally to the borders or to aggregate the HaloDots that have their intersection points too close.

Another problem is related with the representation of the center of the aggregated HaloDot. We have explored two options: the center being the midpoint of the objects represented or the most relevant object. While the first may be more intuitive, the second guides the user's attention to the most relevant objects of his/her search, while still not hiding information about the others.

4 User Study

We have conducted a user study to evaluate whether the proposed solutions assist the user in searching for off-screen objects according with their relevance. We intend to explore and validate our current solutions for the improvement of the representation of the direction, the representation of the relevance and the aggregation to solve the problem of cluttering, as well as to collect users' preferences about our approaches. To compare the HaloDot improvements with the original Halo, we have done a experiment where tasks were presented to participants through scenarios.

We also want to understand the best solution when several design alternatives are explored, namely: aggregation of HaloDots based on the orthogonally cells arrangement or based on closeness of intersection points; always display the number of objects represented by the HaloDot or only if there exists more than one off-screen object; representation of the center of the HaloDot as the most relevant object or as the midpoint of the objects represented.

Considering these objectives, our assumptions in the study were the following:

1. Once we have explained the meaning of colors, we expect that users will not have any difficulties identifying the relevance of PoI. We also expect that users find the most relevant PoI faster that using original Halo.
2. The aggregation is a useful method and the users will prefer the aggregation of HaloDots based on cells arranged orthogonally to the borders, since that means, there are less objects being drawn.
3. The users will prefer the center of the HaloDot being the most relevant object, since that object satisfies their information need.
4. The users will prefer to always see the textual information of the HaloDot, because they will find it easier to work with it.

4.1 Participants and Apparatus

This study had the participation of sixteen volunteer subjects balanced in gender (7 female, 9 male) and background (10 from computer science and 6 from other scientific disciplines and humanities). Their age ranged from 21 to 51 (M=27, SD=9,2). All users had some familiarity with mapping applications on the Web, such as Google Maps for planning routes or finding PoI. Seven out of 16 users had used mobile maps applications occasionally and one uses it daily for finding PoI and navigating in city environments. Participants had experience with zooming and panning but none were familiar with halos. Three participants stated they had used off-screen techniques in video games, such as arrows and mini-map.

The study was carried out on a touch-screen HTC Desire device, running the Android OS 2.2 featuring a 1GHz processor and a 3.7-in touch screen with 480x800 of resolution. During the study, the map covered all the available screen area of the device, and it was based on the Google Maps application.

4.2 Tasks

The study consisted of five tasks that correspond to different scenarios. Before the test the users were informed about the three visualizations techniques in a tourist map application: no contextual clue about off-screen objects; Halo visualization technique and HaloDot. The tasks were performed in the same order (task 1, 2, 3) but the order of the techniques has been counterbalanced as well as the location and relevance configuration of off-screen PoI to reduce sequence and learning effects.

In the first task the participants were asked to find all the PoI (3 restaurants) ordering them in increasing distance from the map center (user position). To execute the task the user had to provide their answer by tapping on the corresponding icon of the PoI. The dependable variable was the time to complete the task and a rating of the difficulty in a five point scale. In addition, at the end of the session the users were also interviewed to rate their preference for a particular visualization and if they understood the characteristics of HaloDot, namely, its textual representation.

In the second task the participants were asked to find the most relevant PoI (i.e. PoI with the highest relevance value), namely a specific restaurant (e.g. “Good Pizza”). The number of PoI is exactly the same (20) for each visualization condition. However, in the HaloDot visualization there is only one red arc that corresponds to the most relevant PoI. The dependable variable was the time to complete the task and a rating of the difficulty in a five point scale. In addition, at the end of the session the users were also interviewed to rate their preference for a particular visualization and if they understood the characteristics of HaloDot, namely, the color as the representation of the relevance, and if the most relevant HaloDot arc (red) is the one that guides their attention.

In the third task we intend to compare the use of HaloDot with two different conditions in the number of relevant PoI: in the first condition, the number of PoI with higher relevance (red arc) is one in a total of 20; in the second one: the number of PoI with higher relevance is higher than one.

The fourth task intends to test the effectiveness of the proposed aggregation approach. The users were requested to explore the map freely, three times, with 10, 50 and 124 PoI, respectively. The last task intends to compare the different design alternatives of HaloDot, namely, presentation of the center of the aggregated HaloDot; the presence of the number of PoI awareness and the aggregation method preferred. The participants were interviewed to rate their preference for a particular option.

4.3 Results

The values of mean completion times for the first task are reported in Fig.4 (left). We have not noticed any relevant result in this task as there is no different effect of the type of visualization configuration (Halo or HaloDot) in the completion time, beyond the difference between the use of an off-screen visualization and no technique at all. At this point few users have noticed the difference between the Halo and the HaloDot, namely in the point of intersection and only one user did not understand the meaning of the number awareness in the HaloDot technique, probably due the existence of few PoI.

For the second task (Fig. 4 middle) a significant main effect was observed for the “type of visualization”. A significant difference was found between Halo and HaloDot. This confirms our first hypothesis: users find the most relevant PoI faster that using original Halo. For the third task and based on Fig. 4 (right) we can see that even with more “relevant HaloDots”, the users were able to more quickly find the desired PoI. This also supports our first hypothesis: once we explained the meaning of colors, we expected that users would not have any difficulties identifying the relevance of PoI. Comments collected after the execution of these tasks revealed that, although few users (31%) have associated the color clue with the relevance of the object, a large number mentioned that the red HaloDots were more visible and provided clues to guide their attention.

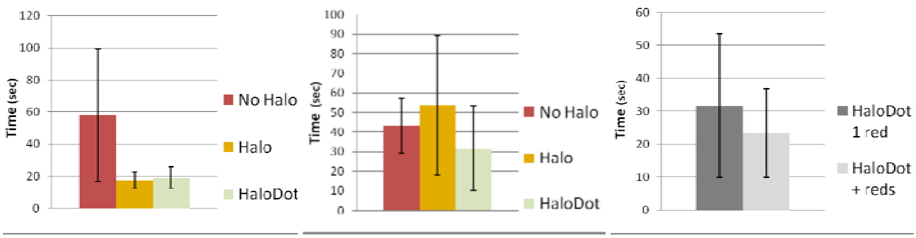


Fig. 4. Mean completion times with standard deviation bars: first task (left); second task (middle); and third task (right)

In the last two tasks, the user expressed their preference between several alternatives of design. In the fourth task, despite the increase of PoI on the map, a large number of users expressed that HaloDot was still usable and that the aggregation was very usefull and 75% prefered the aggregation based on cells arranged orthogonally to the borders. This confirms our second hypothesis. When asked about their preferences about the various configurations of the HaloDot, the results express a positive evaluation of our third and fourth hypotheses: 75% prefered to see the HaloDot with a center on the most relevant PoI, 62% prefered to always see the textual information (number of PoI). The color of the number revealed to be confusing since it was interpreted as the number of the most relevant PoI, instead of being the total number of PoI represented by the aggregate HaloDot, where at least one had a high relevance.

Overall users prefered the HaloDot in comparison with the original Halo, and the main reason pointed was the usage of different colors and the aggregation as they reduce the amount of arcs drawn and consequently improve the perception of the most relevant information. With respect of the utility of the point of intersection as an improvement of direction awarness, users have commented that it was a very helpful feature, however some suggested that other shapes could be used, like arrows.

5 Conclusions and Future Work

In this paper we described the HaloDot technique to improve cluttering problems and to give relevance clues about the off-screen objects. An Android tourist mobile application was developed to compare the proposed technique with Halo. The study showed that HaloDot enables users to search faster for relevant PoI than Halo and that aggregation is a useful method. The next step of this work is to perform a more extended and precise comparison with other variations of the off-screen objects representations (e.g. arrows, lines) and relevance hints. This will enable us to understand and how to optimize them to represent relevance of off-screen objects.

References

1. Baudisch, P., Rosenholtz, R.E.: Halo: a technique for visualizing off-screen objects. In: Proceedings of CHI 2003, pp. 481–488 (2003)
2. Burigat, S., Chittaro, L., Gabrielli, S.: Visualizing location of off-screen objects on mobile devices: A comparative evaluation of three approaches. In: Proceedings of MobileHCI 2006, pp. 239–246 (2006)

3. Zellweger, P.T., Mackinlay, J.D., Good, L., Stefik, M.J., Baudisch, P.: City Lights: contextual views in minimal space. In: Extended Abstracts CHI 2003, pp. 838–839 (2003)
4. Gustafson, S.G., Irani, P.P.: Comparing visualizations for tracking off-screen moving targets. In: Extended Abstracts of CHI 2007, pp. 2399–2404 (2007)
5. Gustafson, S., Baudisch, P., Gutwin, C., Irani, P.: Wedge: clutter-free visualization of off-screen locations. In: Proceeding of CHI 2008, pp. 787–796 (2008)
6. Swienty, O., Reichenbacher, T., Reppermund, S., Zihl, J.: The role of relevance and cognition in attention-guiding geovisualisation. The Cartographic Journal 45(3), 227–238 (2008)
7. Pombinho, P., Carmo, M.B., Afonso, A.P.: Evaluation of Overcluttering Prevention Techniques for Mobile Devices. In: Proceedings IV 2009, pp. 127–134 (2009)
8. Burigat, S., Chittaro, L.: Visualizing references to off-screen content on mobile devices: A comparasion of Arrows, Wedge and Overview + Detail. Interacting with Computers 23, 156–166 (2011)
9. Gonçalves, T., Afonso, A.P., Carmo, M.B., Pombinho, P.: HaloDot: Visualization of the Relevance of Off-Screen Objects. In: SIACG 2011, pp. 117–120 (2011)
10. Reichenbacher, T.: The concept of relevance in mobile maps. Location Based Services and TeleCarthography, 231–246 (2007)
11. Silva, S., Santos, B.S., Madeira, J.: Using Color in Visualization: A Survey. Computer & Graphics 35(2), 320–333 (2011)
12. Wolfe, J.M., Horowitz, T.S.: What attributes guide the deployment of visual attention and how do they do it? Nature reviews. Neuroscience 5, 1–7 (2004)

Using Card Sorts for Understanding Website Information Architectures: Technological, Methodological and Cultural Issues

Helen Petrie, Christopher Power, Paul Cairns, and Cagla Seneler

Department of Computer Science, University of York
{Helen.Petrie,cpower,pcairns,caglaos}@cs.york.ac.uk

Abstract. The card sort technique has many uses in HCI research and practice. Card sorts have traditionally been conducted with physical cards but now programs are available for this task. It is unclear if results from an online version of this technique are as reliable as the "oncard" version. This paper presents a study comparing oncard and online versions of the card sort technique for card set reflecting the information architecture (IA) of two website domains (museum and news sites). No differences were found between the two versions. However, the online version took significantly longer for participants than the oncard version, particularly for non-native English speakers. The card sort technique was also able to reveal cultural differences between mental models of British, Chinese and Indian participants of the IAs of both museum and news websites and showed that all participants have mental models that differ substantially from the typical IAs of websites in these domains.

Keywords: card sort, online card sort program, evaluation methodology, information architecture, website design, museum websites, news websites, cultural differences.

1 Introduction

The card sort technique has many uses in HCI research and practice [6, 15]. In its simplest form it involves asking participants to sort a set of cards with words or pictures on them into groups of cards that are similar. Participants may be asked to provide labels for the groups of cards they have created or they may be provided with pre-defined labels and asked to match the cards to these labels. Card sorts in which participants define the groups are termed open card sorts, those in which the investigator provides the groups are terms closed card sorts [16].

Card sorting is a simple way of gaining insight into the participants' categorizations and mental models of the items represented on the cards. For example, it can assist in the development of the information architecture (IA) of websites, if participants are given cards with the different topics the website is to cover [10].

Although the technique is simple to administer to individual participants, the preparation of cards and the subsequent analysis of the data can be quite tedious and

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 309–322, 2011.

time consuming. It is also easy to make mistakes (as we have found to our cost during this research), particularly if the number of cards in the sort is large. Cards can be mis-numbered, mis-placed or not recorded after the sort, resulting in loss of data.

Card sorts have traditionally been conducted with physical cards but now programs are available (often via the Web) to conduct card sorts virtually, for example on a webpage. This potentially saves time and effort for both participants and investigators [4, 19] and cuts down on the potential for mistakes. We will refer to the two methods as “oncard” and “online” card sorts. Although online card sorts clearly have advantages, do online card sort programs provide the same experience for participants? Anecdotal evidence suggests that online programs (WebSort, XSort, UXSort, Optimal Sort), although they are all relatively easy and clear to use, do not provide quite the same participant experience of being able to move the cards around, look at a card in comparison with one group, then in comparison with another group, and so on. The interaction with tangible objects such as cards is something very natural, which has not yet been completely replaced by virtual experiences. However, does this difference have any effect on the results of card sort studies?

Two studies have investigated the difference between oncard and online card sorts. Harper et al [7] developed their own card sort program and then evaluated it by comparing the program with traditional oncard sorts. 108 participants undertook two open card sorts each, either oncard – oncard, online – oncard, oncard – online, or online – online. Only one set of 33 card items was used, so participants undertook the second card sort with the same card set after a 15 minute distractor task. Although no significant differences were found between oncard and online versions, the fact that participants repeated the sort after only 15 minutes suggests that the results of the second sort would have been contaminated by the first sort.

Bussolon et al [3] also compared oncard and online card sorts, using Netsorting (http://www.cardsorting.info/) for the online sorts. A between participants design was used, with each participant undertaking two card sorts either both online or both oncard. Data from 60 participants were analysed. One set of cards for sorting was of animals (mammals, reptiles, fish, birds) which seems a very easy sort to undertake; the other set was of “gifts”, but no information is provided about the items in this card set. This study used a closed card sort, which is much less common than the open sort and thus the results are less useful. Again no significant differences were found between the oncard and online sorts, but this study is not fully satisfactory in its comparison of the two versions of the technique.

A further question of interest is how many participants are needed to establish a statistically reliable card sort and whether this differs between oncard and online sorts. Tullis and Wood [17] had 168 participants undertake a card sort of 46 cards with an online program (WebSort, http://websort.net/). They then compared the card sort results of 10 samples each of size 2, 5, 8, 12, 15, 20, 30, 40, 50, 60, and 70 participants with the card sort results of the full participant set. They found that with 30 participants there was a 95% correlation with the full card sort and increasing the number of participants beyond 30 provided little increase in the size of the correlation coefficient.

Both Bussolon et al [3] and Tullis and Wood [17] use correlations between the similarity matrices between the cards in two sorts to establish the relationship between the oncard and online sorts and the number of participants needed for a reliable sort. Unfortunately, data in similarity matrices from card sorts is very likely to violate the assumptions for the correlation. In particular, a correlation between variables X and Y assumes that the variance of Y for each value of X is constant, or at least similar (assumption of homogeneity of variance in arrays) [9]. This is not the case for correlations between two card sorts, as some cards always get grouped together and others never get grouped together, leading to a gradation in variance. This violation may not have a substantial effect on the correlation between card sorts, but in our experience obscured small differences between sorts. In this paper we explore the use of the *minimum edit distance* [5] to compare card sorts.

A final question of interest is whether card sorts can help highlight cultural differences in the mental models of participants, particularly in relation to the IA of websites. There is a small but growing body of research and guidelines on how to vary websites for users from different cultures [2, 7], although the guidelines do not mention variations in information architecture. Kralisch, Yeu and Jali [11] used card sorts to investigate the differences between British, German, Malaysian and Russian participants in their understanding of medical terms that might be used in health information websites. They found numerous differences between the four cultural/linguistic groups. Qu et al [14] used card sorts to investigate the differences between the groupings of wedding related images by Chinese and Danish participants. Wan Abdul Rahim, Noor and Aidid [18] were also interested in cultural differences in website information architecture, although they used a questionnaire based on their own and Hofstede's [8] theories of cultural differences to explore Muslim participants' preferences for different information architectures. In our own research [13], we have found that English and Chinese web users preferred different navigation layouts on websites, but it is not clear whether this effect would extend to other aspects of IA.

We therefore conducted a study comparing oncard and online card sorting methods with two different card sorts relating to IAs of websites to address the following questions:

- are the cards sorts produced oncard and online any different from each other?
- is the participant experience equivalent in undertaking a card sort exercise online and with physical cards?
- is an online card sort more efficient for both researchers and participants than an oncard card sort?
- can card sorts reveal cultural differences in participants' categorizations and mental models of the IA of two types of websites?

2 Method

In order to investigate these questions, a two by two non-factorial between participants study was undertaken. Each participant undertook two different card sorts, one online and one oncard, with one being about the IA of news websites and the other being about the IA of museum websites.

2.1 Participants

A total of 218 participants took part in the study, comprising students from two undergraduate modules and one graduate module on interactive systems in the Department of Computer Science at the University of York. There were 188 male and 30 female participants, aged between 18 to 35 years, with a mean age of 21.3 years. Within the participant group there were 116 who were native speakers of English, 102 were native speakers of a variety of other languages.

Of the 102 students who spoke English as a second language the average self-reported proficiency in English was 5.4 on a 7 point scale, with 7 being "Very Proficient" to 1 "Not at all Proficient". The students came from a variety of non-English speaking cultural backgrounds including: Chinese, Taiwanese, Indian, other European cultural groups including Polish, German and Greek, Middle-Eastern, and sub-Saharan Africa.

Participants undertook the study as part of the practical assignments for their module. However, to motivate student participation, there was a lottery draw for three £10 gift vouchers for a major online book retailer for those who completed the assignment.

2.2 Equipment and Materials

For creation of the word sets to be used in the card sorts, a survey was undertaken of 10 news websites and 18 museum websites from countries with a national language of English (Australia, Canada, UK, USA). For each website the top level IA menu was recorded. From the words within these menus the top 7 most commonly occurring words were chosen as base categories for collecting the sets of words for each website domain.

Each website was visited again and the IA under each of the 7 categories was recorded and grouped together according to their meaning. For example, if material on personal finance was labelled as "Finances" on one news website while on another it was "Your money", these were recorded under one group "Personal Finance" and as occurring on 2 distinct websites. When complete, the top groups that occurred in the majority of the websites under the 7 categories were selected as words for the card sort. The news sites produced a set of 50 words while museums produced a set of 40 words fitting these criteria. These groups are referred to subsequently as the "a priori" groups.

Within each set of words, each word was given a distinct number. The words were printed onto sets of cards sized 89mm x 51mm and each word's associated number was recorded on the back of the card.

The online card sort was prepared the online card sorting package WebSort (http://www.websort.net). Two different online card sorts, one for news websites and one for museum websites, were created. Each word was entered into the WebSort software in the same order as the number assigned to the cards to ensure that cards could be matched during analysis. The online sort was set to request a participant number to allow matching of online sorts with the respective oncard sorts for each participant. The online sort was set such that it would randomize the order of presentation of the cards for each participant and so that participant results would only be recorded if all cards were sorted and the categories into which they were sorted were labelled.

A questionnaire regarding the participants' cultural background, language proficiency and other demographic information was prepared in the online survey package QuestionPro (www.questionpro.com).

All computer based tasks were undertaken in computer labs within the Department of Computer Science at the University of York, with participants undertaking the tasks in their preferred web browser (e.g. Firefox, Internet Explorer, Opera) and their preferred operating system (e.g. Windows, Linux).

Tickets with participant numbers and a sheet recording which order of conditions each number should undertake were prepared for the participants. The order of conditions for participants were created through a complete counterbalancing of the two conditions, online versus oncard and museum versus news. After each session the counterbalancing was adjusted by the researchers to account for missing data points from participants who failed to complete all the components of the study.

2.3 Procedure

Data collection sessions took place during the practical class periods of each module that participated. On arrival at the class each participant was given a lottery ticket to serve as their participant number during the session (and as their number in the prize draw) and a sheet of instructions. The practical leaders and the researchers reviewed the instructions verbally with the participants at the beginning of the session.

The first step of the study was for the participant to fill out the questionnaire regarding their cultural background and demographic information. This questionnaire included a declaration of informed consent that included a statement that the participant could withdraw at any time without academic penalty.

Each participant was instructed to undertake two sorts, one sort on the news words and one sort on the museum words. Participants were then asked to check the counterbalance sheet for order in which they were to conduct these sorts and which sort should be undertaken online and which should be undertaken oncard.

2.3.1 Oncard Procedure

For oncard sorts participants were handed an envelope containing a set of cards for the condition (news or museum) they were undertaking. Participants were also given a set of blank slips of paper on which to record their category labels.

Participants started the oncard sort with the cards in a stack with words facing up. They were asked to shuffle the cards at the beginning of their sort to ensure as much randomness as possible in the placement of cards in the set. The participants were told that they were to sort the cards into categories of their choosing and label those categories with the blank sheets of paper provided. They were told that there were no correct answers, but that they should have more than one category and fewer than the maximum number of cards. They were informed that they were not to try for the fastest time possible, but that they should also not go too slowly so as to avoid indecision about categories.

The participants were timed by another member of the class for how long their sort took from the beginning of the sort to the last category being labelled.

When the oncard sort was complete, the participant recorded the label of each category in a spreadsheet provided and then the number of the cards in each pile and the cards associated with that pile. The participants also recorded their participant lottery number and the time it took to undertake the sort in the spreadsheet.

2.2.2 Online Procedure

For the online sort participants were instructed to go to the web address for the sort they were assigned to undertake. On arriving at the website the participant was presented with a dialog box requesting their participant number. After this, the website presented the same instructions as those described for the oncard sort.

At the end of the sort, the participants' labelled categories, the pile names and their time performing the sort were automatically recorded by the website.

2.3 Data Preparation and Analysis

Due to the need to combine a variety of sources of data for participants, specifically demographic data, online sort data and oncard sort data, a bespoke piece of software was written to aggregate and clean the data for each participant. This software was developed using a test-driven development methodology [1] with each iteration of development identifying errors in the data that could be automatically identified and removed. After applying the final software suite to the data, participants who had any of the following errors in their data (all of which occurred), usually the result of input errors in the oncard spreadsheets, were removed from the datasets:

- sorts where the participant number and demographics were missing from the questionnaire results;
- sorts with cards missing in the recording of either card set;
- sorts with duplicate cards recordings;
- sorts where the number of cards recorded in each pile was not equal to the actual number of cards in the pile;
- sorts where category names were repeated.

For each card sort data set a hierarchical cluster analysis [9] was conducted to extract the average grouping of the cards by the relevant set of participants, these groupings will be termed *card groups*. This analysis uses a similarity matrix of the number of times each card in the set is grouped with each other card in the set.

In order to compare two different card sorts this paper uses the *minimum edit distance* [5]. This *metric*, which is symmetric, positive and satisfies the triangle inequality, indicates the minimum number of card changes, from one group in a sort to another group in the same sort, in order to transform one sort result into another sort result. This metric is polynomial in time complexity, $O(n^3)$, where n is the number of cards in the card set, and can be calculated by hand easily on card sorts with small numbers of cards. In this paper the minimum edit distance will be expressed as the percentage of the total cards in the card set, in order to allow comparisons of the results between two card sets of different sizes.

3 Results

To investigate the question of whether the cards sorts conducted oncard were different from those conducted online, minimum edit distances between the cards sorts from all the native English speakers were compared (English speakers were used in this analysis simply to create a more homogenous data set). The results for the oncard and online sorts were compared with each other and also compared with the a priori groups. The results are shown in Tables 1 and 2.

The first result of note is that for both card sets, but particularly for the news card set, both the oncard and online sorts produced groups that were substantially different from the a priori card set, although the latter reflects the average organization of items on a set of real web sites. In the case of the museum card set, 27.5% of cards were grouped differently in the oncard sort compared to the a priori groups and 22.5% of cards were grouped differently in the online sort compared to the a priori card set. The situation was even more extreme with the news card set, with 36.0% of cards grouped differently for both the oncard and online sort compared to the a priori groups. This suggests that young English speaking participants (almost all of whom were British) do not have the same mental models of museum and news categories as those who develop websites in these areas.

Tables 3 and 4 show the differences in groups between the a priori, oncard and online sorts for museum and news card sets respectively. In the museum set, it can be seen that there were a number of differences in which cards were placed in particular groups, but in particular, in both the oncard and online sort, the participants produced an additional group ("Family activities'), that did not occur in the a priori groups. In the news set, participants produced fewer groupings than the a priori set. The "Business" and "Lifestyle" groups did not appear at all in the participants' sorts. In addition, the classic news categories of factual and opinion pieces were very blurred in the participants' sorts.

Table 1. Minimum edit distance (%) between a priori, oncard, and online card sorts for museum card set for native English speakers

	A priori	Oncard (N = 46)	Online (N = 54)
A priori	-	27.5	22.5
Oncard	27.5	-	7.5
Online	22.5	7.5	-

Table 2. Minimum edit distance (%) between a priori, oncard, and online card sorts for news card set for native English speakers

	A priori	Oncard (N = 30)	Online (N = 52)
A priori	-	36.0	36.0
Oncard	36.0	-	6.0
Online	36.0	6.0	-

Table 3. Museum card set: a priori groups, oncard and online sorts

A priori groups	Oncard groups	Online groups
Exhibitions	**Exhibitions**	**Exhibitions**
Current exhibitions	Current exhibitions	Current exhibitions
Future exhibitions	Future exhibitions	Future exhibitions
Past exhibitions	Past exhibitions	Past exhibitions
Travelling exhibitions	Travelling exhibitions	Travelling exhibitions
The Collection	**The Collection**	**The Collection**
Our collections	Our collections	Our collections
Search the collection	Search the collection	Search the collection
Conservation	Online collection	Online collection
Online collection	Highlights of the collection	Collections management
Collections management		Highlights of the collection
Highlights of the collection		
The Shop	**Shopping and Eating**	**Shopping and Eating**
Books and media	Books and media	Books and media
Prints and posters	Prints and posters	Prints and posters
Jewellery	Jewellery	Jewellery
Fashion and accessories	Fashion and accessories	Fashion and accessories
Homewares	Homewares	Homewares
Stationery	Stationery	Stationery
	Eat and drink	Eat and drink
	Shops	Shops
Learning	**Learning**	**Learning**
Schools	Schools	Schools
Teachers	Teachers	Teachers
Online resources	Adult learners	Online resources
Adult learners	Learning centre	Adult learners
Activities for families	Talks and lectures	Learning centre
Learning centre	Courses and demonstrations	Talks and lectures
		Courses and demonstrations
		Events calendar
Visiting the museum	**Visiting the museum**	**Visiting the museum**
Finding the museum	Finding the museum	Finding the museum
Opening times	Opening times	Opening times
Eat and drink	Access for disabled visitors	Access for disabled visitors
Access for disabled visitors	Booking tickets	Booking tickets
Booking tickets	Events calendar	Contact us
Shop	Jobs	
Family visits		
The Organization	**The Organization**	**The Organization**
Jobs	Jobs	Jobs
Contact us	Contact us	Press Room
Press Room	Press Room	Management
Management	Management	History of the Museum
History of the Museum	History of the Museum	Mission statement
Mission statement	Mission statement	Volunteering at the museum
Volunteering at the museum	Volunteering at the museum	Conservation
	Online resources	
	Conservation	
	Collections Management	
	Family activities	**Family activities**
	Events for the family	Events for the family
	Activities for families	Activities for families
	Family visits	Family visits

Table 4. News card set: a priori groups, oncard and online sorts

A priori groups	Oncard groups	Online groups
Arts	**Arts**	**Arts**
Films	Comics	Books
Books	Films	Comics
Music	Health	Films
Stage and Dance	Markets	Music
TV & Radio	Music	News Discussions
Comics	News Discussions	Relationships
Visual Arts	Relationships	Stage and Dance
Markets	Stage and Dance	TV & Radio
	TV & Radio	Visual Arts
	Visual Arts	
Business		
Economics		
Careers		
Small Business		
Industries		
Personal Finance		
Lifestyle		
Food and Drink		
Fashion and Style		
Health		
Family		
Homes		
Relationships		
Money	**Finance**	**Finance**
Savings	Books	Borrowing
Property	Borrowing	Careers
Taxes	Careers	Economics
Investments	Economics	Family
Pensions	Industries	Industries
Borrowing	Insurance	Insurance
Insurance	Investments	Investments
	Pensions	Markets
	Personal Finance	Pensions
	Property	Personal Finance
	Savings	Property
	Small Business	Savings
	Taxes	Small Business
		Taxes
News	**Factual**	**Factual**
National	Basketball	Basketball
World	Columnists	Columnists
Politics	Education	Education
Education	Family	Fashion and Style
Science	Fashion and Style	Health
Local	Homes	Homes
Technology	Local	Local
	National	National
	Politics	Politics

Table 4. (*Continued*)

	Factual (cont) Puzzles and Games Science World	**Factual (cont)** Puzzles and Games Science World
Opinion Columnists Letters to the Editor Blogs Cartoons News Discussions Editorials Corrections Leading articles Commentators	**Non-factual** Blogs Cartoons Commentators Corrections Editorials Leading articles Letters to the Editor Technology	**Non-factual** Blogs Cartoons Commentators Corrections Editorials Leading articles Letters to the Editor Technology
Sport Football Golf Tennis Motor Sport Ice Hockey Baseball Basketball	**Sport** Baseball Food and Drink Football Golf Ice Hockey Motor Sport Tennis	**Sport** Baseball Food and Drink Football Golf Ice Hockey Motor Sport Tennis

However, the oncard and online card sorts produced very similar groups, with less than 10% of cards grouped differently in either sort (7.5% for the museum set, 6.0% for the news set). This shows that the oncard and online procedures produce equivalent sorts, so the different technology does not affect the outcome. It also supports the finding that the groups produced by participants are different from the a priori groups, as both the oncard and online produced the same set of groups, although different from the a priori groups.

To investigate one aspect of the user experience of undertaking a card sort, and also to investigate the efficiency of oncard and online sorts, the times taken to complete the card sorts using the two versions of the technique were analysed. For this analysis, data from the native and non-native speakers of English were considered separately, as it was thought that the non-native speakers may take longer to do the card sorts both oncard and online, as they would need to think about the words more. Two independent measures two way analyses of variance were calculated, one each for the museum and news sorts, on the time taken to complete the sort with the following variables: Sort version (oncard vs online) and Language (native English speaker vs non-native English speaker). For both sorts there was a significant difference between the sort versions (for museum: $F = 12.92$, $df = 1, 182$, $p < 0.001$; for news: $F = 6.60$, $df = 1,153$, $p < 0.01$), with oncard sorts producing significantly shorter times than online sorts. There was also a significant difference between native and non-native speakers of English, as predicted (for museum: $F = 43.75$, $df = 1, 182$, $p < 0.001$; for news: $F = 25.48$, $df = 1,153$, $p < 0.001$), with native speakers producing significantly shorter times than non-native speakers. Finally there was a significant interaction between the effects of sort version and user group (for museum: F = 6.74, df = 1, 182, $p < 0.01$; for news: F = 25.48, df = 1,153, $p < 0.001$). These interactions are illustrated in Figures 1 and 2, which show that for the non-native speakers of English, the time increase for online compared to oncard is magnified.

To investigate whether card sorts would show cultural differences in the mental models of participants of the IAs for the two website domains, card sorts from participants from the following cultural groups were compared: British (we use British to indicate the cultural group, English to indicate the language group), Chinese and those from the Indian sub-continent. The results are shown in Table 5 and 6. The Indian sorts are even more different from the a priori than the British sorts (40.0% for museums and 42.0% for news). For the Chinese sort, the museum sort is closer to the a priori groups than the British (although the number of participants in this group is only 16, so we have less confidence in this sort than the others, for which the numbers are higher), although the news sort is less similar to the a priori than the British (38.0%).

For the museum set, the number and general nature of groups produced by the British, Indian and Chinese sorts are the same. However, the particular cards grouped together are rather different for some of the groups. In particular, the cards in the groups "The organization" and "Visiting the museum" differ considerably (space does not permit us to reproduce the full groups in this analysis, but they are available from the authors). The Indian sort produced a larger group for "Visiting the museum" with a number of cards from the a priori group "The organization" included in it. Conversely, the Chinese sort included more cards in the "The organization" rather than in "Visiting the museum".

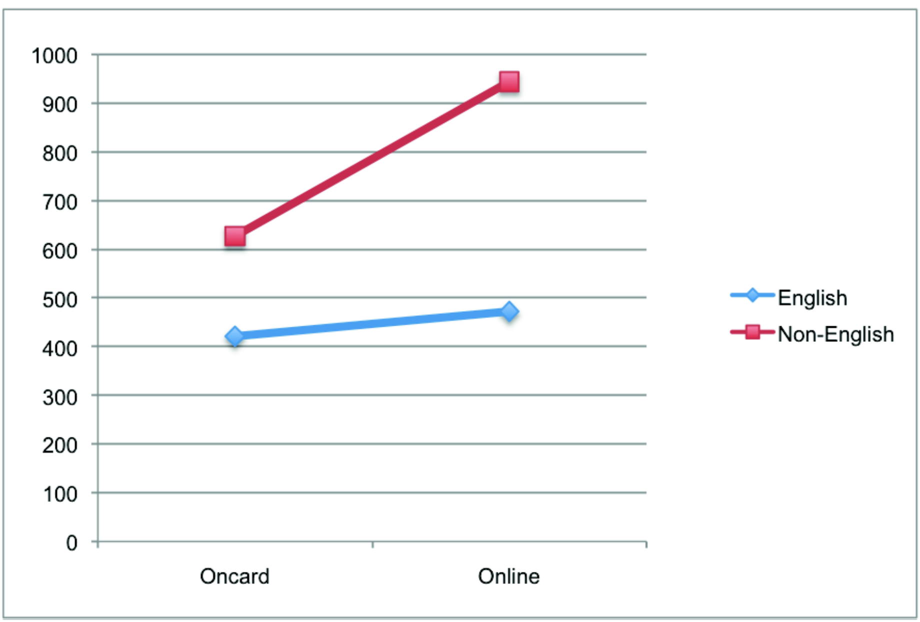


Fig. 1. Mean time (seconds) to complete museum card sort oncard and online for native English speakers and non-native English speakers

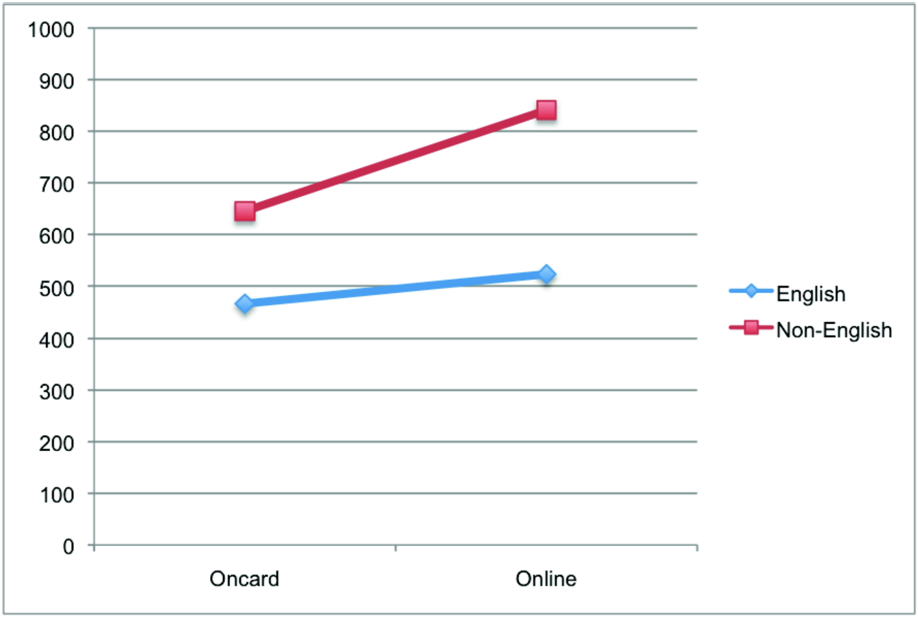


Fig. 2. Mean time (seconds) to complete news card sort oncard and online for native English speakers and non-native English speakers

Table 5. Minimum edit distance between card sorts by British, Chinese, and Indian participants for museum card set

	A priori	British (N = 100)	Indian (N = 24)	Chinese (N = 16)
A priori	-	25.0	40.0	22.5
British	25.0	-	15.0	15.0
Indian	40.0	15.0	-	30.0
Chinese	22.5	15.0	30.0	-

Table 6. Minimum edit distance between card sorts by British, Chinese, and Indian participants for news card set

	A priori	British (N = 82)	Indian (N = 21)	Chinese (N = 20)
A priori	-	34.0	42.0	38.0
British	34.0	-	14.0	18.0
Indian	42.0	14.0	-	16.0
Chinese	38.0	18.0	16.0	-

For the news set, the number and nature of groups produced by the three cultural sorts varied. The British sort produced six groups, whereas the Indian and Chinese sorts produced only five groups each. All sorts produced similar groups for "Money and business", "Sports" and "Factual news". However, the British sort produced groups for "Opinion", "Arts" and "Non-factual", whereas the Indian and Chinese sorts produced "Entertainment" and "Non-factual" with rather different groupings of particular cards.

4 Discussion and Conclusions

This study has shown a number of interesting results about the card sort technique and its use. In spite of anecdotal evidence, we found no difference between conducting card sorts on physical cards and using an online program (oncard versus online) with respect to the reliability of the results. Less than 10% of cards were grouped differently and the groups which emerged were the same. This finding is in line with previous studies by Bussolon et al [3] and Harper et al [7]. However, both oncard and online sorts produced groups from English speaking participants which were substantially different from the a priori groups, although the latter were based on an analysis of a substantial number of websites in each of the two domains. This suggests that website designers in these domains might do well to study their audiences more carefully, and possibly use card sort studies to elicit the mental models of these audiences.

A further unexpected result emerged from the times taken for participants to complete card sorts that were significantly longer using the online program than with physical cards. This effect was significantly more pronounced for non-native speakers of English. Thus the online version of the card sort technique may well save time and effort for researchers and practitioners but the time for each participant is significantly longer than with physical cards. Researchers and practitioners need to consider the balance of importance between these two factors.

Finally, the study showed that the card sort technique can be used successfully to reveal cultural differences in the mental models of information architectures of website domains. Interesting and meaningful differences were found between British, Chinese and Indian participants in their average groupings of cards relating to both museum and news website information architectures.

In further analyses, we will investigate the number of participants required to establish a stable card sort using the minimum edit distance statistic. This may produce different results from those proposed by Bussolon et al [3] and Tullis and Wood [17], which were based on correlation analyses.

This study has confirmed a number of interesting and useful facts about the card sort technique and its usefulness in the HCI domain. In addition, it has revealed that this seemingly simple tool can be used to draw out subtle and meaningful information about users and their mental models. Thus, it is important to expose new researchers and practitioners in its application in the design of interactive systems.

References

1 Astels, D.: Test Driven Development: A Practical Guide. Prentice Hall Professional Technical Reference (2003)

2 Aykin, N., Quaet-Faslem, P.H., Milewski, A.E.: Cultural ergonomics. In: Salvendy, G. (ed.) Handbook of Human Factors and Ergonomics, 3rd edn. John Wiley and Son, New Jersey (2006)

3 Bussolon, S., Russi, B., del Missier, F.: Online card sorting: as good as the paper version. In: Proceedings of the 13th European Conference on Cognitive Ergonomics: Trust and Control in Complex Socio-Technical System (ECCE 2006). ACM Press, New York (2006)
4 Chaparro, B.S., Hinkle, V.D., Riley, S.K.: The usability of computerized card sorting: a comparison of three applications by researchers and end users. Journal of Usability Studies 4(1), 31–48 (2008)
5 Deibel, K., Anderson, R., Anderson, R.: Using edit distance to analyze card sorts. Expert Systems 22(3), 129–138 (2005)
6 Fincher, S., Tenenberg, J.: Making sense of card sorting data. Expert Systems 22(3), 89–93 (2005)
7 Harper, M.E., Jentsch, F., van Duyne, L.R., Smith-Jentsch, K., Sanchez, A.D.: Computerized card sort training tool: is it comparable to manual card sorting? In: Proceedings of the Human Factors and Ergonomics Society 46th Annul Meeting, pp. 2049–2053 (2002)
8 Hofstede, G., Hofstede, G.J.: Cultures and organizations: software of the mind (Revised and expanded 2nd edn.). McGraw-Hill, New York (2005)
9 Howell, D.C.: Statistical methods for psychology, 6th edn. Thomson, Wadsworth (2007)
10 Katsanos, C., Tselios, N., Avouris, N.: AutoCardSorter: Designing the information architecture of a web site using latent semantic analysis. In: Proceedings of the 26th Annual SIGCHI Conference on Human factors in Computing Systems (CHI 2008). ACM Press, New York (2008)
11 Kralisch, A., Yeo, A.W., Jali, N.: Linguistic and cultural differences in information categorization and their impact on website use. In: Proceedings of the 39th Hawaii International Conference on System Sciences (HICSS 2006). IEEE Computer Society, Los Alamitos (2006)
12 Marcus, A.: Global/intercultural user interface design. In: Sears, A., Jacko, J. (eds.) The Human Computer Interaction Handbook, 2nd edn. Lawrence Erlbaum Associates, New York (2003)
13 Petrie, H., Power, C., Song, W.: Internationalization of Websites: Navigation in Chinese and English Language Sites. In: Aykin, N. (ed.) IDGD 2009. LNCS, vol. 5623. Springer, Heidelberg (2009)
14 Qu, W., Sun, X., Nawaz, A., Plocher, T., Clemmensen, T.: Cultural differences between Chinese and Dane in card sorting. In: Proceedings of the 8th Pan-Pacific Conference on Occupational Ergonomics, PPCOE 2007 (2007)
15 Rugg, G., McGeorge, P.: The sorting techniques: a tutorial paper on card sorts, picture sorts and item sorts. Expert Systems 14(2), 80–93 (1997)
16 Tullis, T., Albert, B.: Measuring the user experience. Morgan Kaufmann, San Francisco (2008)
17 Tullis, T., Wood, L.: How many users are enough for a card-sorting study? In: Proceedings of the Usability Professionals Association Conference 2004, Minneapolis, MN, June 7-11 (2004)
18 Wan Abdul Rahim, W.M.I., Noor, N.L.M., Aidid, S.S.S.H.: Culture influences to website information architecture: an empirical investigation. In: Proceedings of Information Technology 2008 (ITSim 2008). IEEE Computer Society, Los Alamitos (2008)
19 Zavod, M.J., Rickert, D.E., Brown, S.H.: The automated card-sort as an interface design tool: a comparison of products. In: Proceedings of the Human Factors and Ergonomics Society 46th Annul Meeting, pp. 646–650 (2002)

The Treatment of Temporal Data in Web-Based Reservation Systems: An Inspection-Based Evaluation

Gerhard F. Knolmayer, Lukas E. Helfenstein, and Viola Sini

University of Bern,
Institute of Information Systems,
Engehaldenstrasse 8, CH-3012 Bern,
Switzerland
{gerhard.knolmayer,lukas.helfenstein,
viola.sini}@iwi.unibe.ch

Abstract. Web-based reservation systems realize a broad variety of different and often inadequate ways of handling temporal data in their user interfaces. We compiled possible procedures for treating temporal data in reservation systems in a Morphological Box, ending up with 49 treatment options for 15 features. We initiated an usability inspection by 15 usability experts, asking for evaluation of the user-friendliness of these options and the relevance of their differences. After discussing the results we use them to develop an evaluation model. We draw profile lines to compare existing systems, weight the scores of the options by relevance factors of the features, and compute aggregate usability indicators to rank the systems. In our analysis of 60 airlines, Delta Air Lines and American Airlines reach the best results. We also show that companies that belong to the same group differ remarkably in the usability of their reservation systems.

Keywords: Temporal Data, Morphological Analysis, Reservation Systems, Usability, Usability Indicators, Usability Inspection.

1 Introduction

Web-based reservation systems have become a highly relevant distribution channel in the travel industry [1]. Nevertheless, the poor usability of these systems is often criticized [2-8]. Handling temporal data in Web interfaces is an essential feature of reservation systems because each reservation is time related. However, temporal data is treated by Web-based reservation systems quite differently, sometimes in user-unfriendly ways and in contradiction to usability guidelines, and in some cases even erroneously [9].

In this paper we focus on handling temporal data in Web-based reservation systems and, thus, consider a single element of the user interface in more detail than previous analyses. Of course appropriate treatment of temporal data is only one aspect in the evaluation of user interfaces and the usability of reservation systems. However, inade-

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 323–339, 2011.

quate treatment of temporal data seems to carry considerable potential to demotivate a client, may result in aborting a planned reservation, and can result in additional transaction costs or loss of revenues. It is widely accepted that a typical website carries so many usability problems that only a fraction of them will be identified in an empirical analysis; thus it seems legitimate to focus on one special aspect.

The remainder of this paper is organized as follows: In Section 2 we reference the large body of research on treating time in information systems with special emphasis on usability issues. In Section 3 we discuss Morphological Boxes and apply them to establish a systematic approach on discussing the options for treating temporal data in user interfaces. We also give examples for reservation systems which apply the different options. Section 4 describes an inspection-based evaluation of these options by 15 usability experts, discusses some details, and visualizes the results. In Section 5 we develop usability indicators for treating temporal data and determine the associated indicator values for selected reservation systems. Section 6 summarizes the results and gives an outlook.

2 Temporal Data as Research Object

2.1 Temporal Data in General

The adequate treatment of temporal data has been a major research area as well in Computer Science as in Management of Information Systems. There are thousands of papers which are referenced in at least 13 bibliographies on different aspects of handling temporal data in information systems [10, 11]. The most widely covered area is temporal databases; a recently published Encyclopedia of Database Systems contains more than 80 contributions related to temporal data [12, 13]. Temporal Logic and Temporal Reasoning are important research areas in Artificial Intelligence. Only a few papers focus on temporal data in Web-based systems. Furthermore, the potential problems of inadequate treatments of temporal data aroused broad public interest in connection with solving the Year 2000 (Y2K) problem.

2.2 Usability Aspects of Temporal Data

Fabre and Howard [14] argued that good practices for handling temporal data should find their way into style guides, screen design guidelines, and repositories of reusable interface designs; more generally, Temporal Aspects of Usability (TAU) should be given more attention. TAU have been discussed in several workshops, but most definitions of usability at best leave temporal aspects implicit, and at worst omit any consideration of time [14]. Several design guides for date entry and display exist [15-18]. Bainbridge [19] provides detailed recommendations to system developers for handling temporal data in reservation systems. Information engineering concepts like services and patterns could be useful also for treating temporal data. Temporal services are compiled in [20], temporal patterns discussed, e.g., in [21].

3 Morphological Boxes for Treating Temporal Data

A Morphological Analysis supports an ordered way of looking at things [22, 23]. A Morphological Box represents relevant features and the associated options in existing or forthcoming products and services. It can be applied to provide a systematic overview of elementary courses of action and how they are or can be combined.

In [24] we developed a Morphological Box to show the options for handling temporal data in reservation systems. The rows of the Box describe main features of treating temporal data in reservation systems and the corresponding table entries show the options available for each feature. The features were found by literature analysis and by a critical review of many reservation systems and are described in some detail in [24].

The research documented in this paper aimed to evaluate the options described in the Morphological Box by Usability Experts to allow well-founded statements about the usability of different reservation systems and to show their improvement potentials. When we designed the questionnaire we found that some items used in [24] would need refinement. Thus we rearranged the Morphological Box which now contains 15 instead of 10 features and 49 instead of 33 options. The features in Table 1 differ from those in [24] as follows:

- We now distinguish between initial entry of temporal data and changes.
- The feature "Temporal Integrity Constraints" is refined into 3 subcases of keeping temporal integrity (cf. [9]).
- We distinguish setting default values on the initial screen for entering temporal data from setting nontrivial default values after user entry.
- We also differentiate whether temporal flexibility is offered at the entry page or after the search procedure.

We number the features in the first column of Table 1 and the options for each feature in the table elements. Table 1 also gives *examples of Web-based reservation systems* that realize the different options (as of March 2011). In our extensive Web searches we found examples for 47 of the 49 options. This result indicates that existing systems use a broad variety of different procedures for treating temporal data.

When developing a Morphological Box one has to consider that for each solution one and only one marking should appear in each row of the Box. In this case the markings can be connected to obtain a profile line for a certain solution. It does not suffice to present elementary options in the Morphological Box if these may also be combined. For instance, if keyboard, dropdowns, and calendars are the 3 basic options for entry and changes of temporal data, one could assume that the relevant number of combinations is

$$\binom{3}{1} + \binom{3}{2} + \binom{3}{3} = 7$$

But providing a dropdown typically excludes entering dates via the keyboard. Thus, the number of relevant options for Features 3 and 4 in Table 1 is

$$\binom{3}{1} + \binom{3}{2} - 1 = 5$$

However, in our analysis of reservation systems we also found some strange exceptions of this assumption (cf. Section 5.2).

Formally, the upper bound on the number of combinations in Table 1 is almost 25 millions but not all of them are feasible because some entries in Table 1 are related. For instance, the option "No calendar" exists for Features 5 and 6. If "No calendar" is true for Feature 5, only this option is feasible for Feature 6; instead of 4*3=12 only 3*2+1=7 combinations result from joining Features 5 and 6.

4 Evaluating the Options

4.1 Design of the Evaluation

Each Web-based reservation system can be characterized by a profile line which connects the markings in the Morphological Box that are associated with the options realized by the system. To evaluate the treatment of temporal data in a given system one has to appraise the options described in the Morphological Box and also to evaluate the relevance of the features. The latter is similar to classifying usability problems on a scale of severity in a usability inspection [25].

For evaluating the usability of systems one can refer to users or to usability experts. As in other evaluations of information system design, the number of reviewers should be limited to a rather small number of persons. For usability studies, the participation of experts with broad background in usability evaluation and human-computer interface design is recommended. It is often assumed that no more than 5 experts are needed to obtain sufficient results [26, 27].

In our case we had to modify the typical procedure applied in an Inspection-Based Evaluation [28] because we did not want to have one specific solution evaluated but the experts should evaluate all options described in Table 1. Therefore we tried to find a larger number of evaluators than suggested for practical applications. We were able to interest 5 experts working in consultancy and interface development firms and 4 academics who are doing research in connection with evaluating the business benefits of websites to participate in the study. Another 3 evaluators were very familiar with Web-based reservation systems. Three Ph.D. students with a broad range of practical experiences also participated in the study. In total 15 persons evaluated the options described in Table 1. We explained the options in detail and gave, whenever possible, references to existing systems, showing screenshots, and providing links to systems in which this option was implemented.

Table 1. Morphological Box for Treating Temporal Data (Part 1)

Number	Feature	Characteristics / *Examples*				
1	**Definition of Booking Interval**	① t_s, t_e *Most systems*	② t_s, d *www.hotelpronto.com*	③ d, t_e *No system found*	④ t_s, d, t_e *hotels.myswitzerland.com/?idlng=1*	
2	**Entry of Temporal Data**	① Only Keyboard *ww1.shanghai-air.com/salnewweb-en/index.aspx*	② Only click in dropdowns *www.evaair.com/html/b2c/english/*	③ Only click in calendar *www.germanwings.com/en*	④ Keyboard or click in calendar *www.swiss.com*	⑤ Click in dropdown or calendar *www.ana.co.jp/asw/wws/sz/e/*
3	**Changes of Temporal Data**	① Only Keyboard *ww1.shanghai-air.com/salnewweb-en/index.aspx*	② Only click in dropdowns *www.evaair.com/html/b2c/english/*	③ Only click in calendar *www.germanwings.com/en*	④ Keyboard or click in calendar *www.swiss.com*	⑤ Click in dropdown or calendar *www.ana.co.jp/asw/wws/sz/e/*
4	**Display of Date Format**	① Not indicated *www.southwest.com*	② Abstractly (e.g., mm/dd/yy) *www.delta.com*	③ Numerically (e.g., 09/11/11) *www.united.com*	④ With months shown as text *www.qantas.com*	
5	**Display of Calendars**	① No calendar *www.evaair.com/html/b2c/english/*	② After clicking in date entry field *www.swiss.com*	③ After clicking on calendar icon *www.lufthansa.com*	④ Without user action *www.airberlin.com*	
6	**Representation of and Access to Temporal Data**	① No calendar *www.evaair.com/html/b2c/english/*	② Sequential access to months *www.austrian.com*	③ Direct access to months *www.qantas.com*		

Legend:
t_s … Start of reservation interval; d … Duration; t_e … End of reservation interval

Table 1. Morphological Box for Treating Temporal Data (Part 2)

Number	Feature	Characteristics / *Examples*			
7	**Temporal Integrity of Start Date**	① Not supported *www.singaporeair.com*	② Past dates are excluded in dropdowns or calendar *Many systems, e.g., www.swiss.com*		
8	**Temporal Integrity of End Date**	① Not supported *www.ryanair.com/en*	② Infeasible dates are excluded in dropdowns or calendar *Many systems, e.g., www.swiss.com*		
9	**Temporal Integrity of Date Relationships**	① Not supported *www.easyjet.com/asp/en*	② Defining infeasible dates is excluded *www.delta.com*	③ Avoided by setting default values *www.united.com*	
10	**Default Temporal Data on Initial Screen**	① No default values *www.jetblue.com*	② Default values set *Most systems*		
11	**Nontrivial Default Temporal Data after User Entry**	① No nontrivial default values *www.lufthansa.com*	② Nontrivial default values added without warning *www.united.com*	③ Nontrivial default values added with warning *www.easyjet.com/asp/en*	
12	**Error Messages**	① Avoided by setting nontrivial default values *www.united.com*	② Immediate at data entry (e.g., by erasing previous date entries) *www.kulula.com*	③ Delayed until subsequent user action (e.g., Search) *www.singaporeair.com*	
13	**Travel Day Flexibility on Entry Page**	① No flexibility offered *www.austrian.com*	② Flexibility interval(s) >0, but undefined *www.ryanair.com*	③ Flexibility with preset interval(s) (e.g., ± 3 days) *www.lufthansa.com*	④ Flexibility with arbitrarily user-defined interval(s) *No system found*
14	**Travel Day Flexibility after Search**	① No flexibility offered *www.austrian.com*	② Several days shown but only one price per day *www.lufthansa.com*	③ Several days shown with several prices per day *www.airberlin.com*	
15	**Temporal Data after "Major Changes"**	① Temporal data is lost *www.ana.co.jp/asw/wws/sz/e/*	② Temporal data is kept *www.lufthansa.com*		

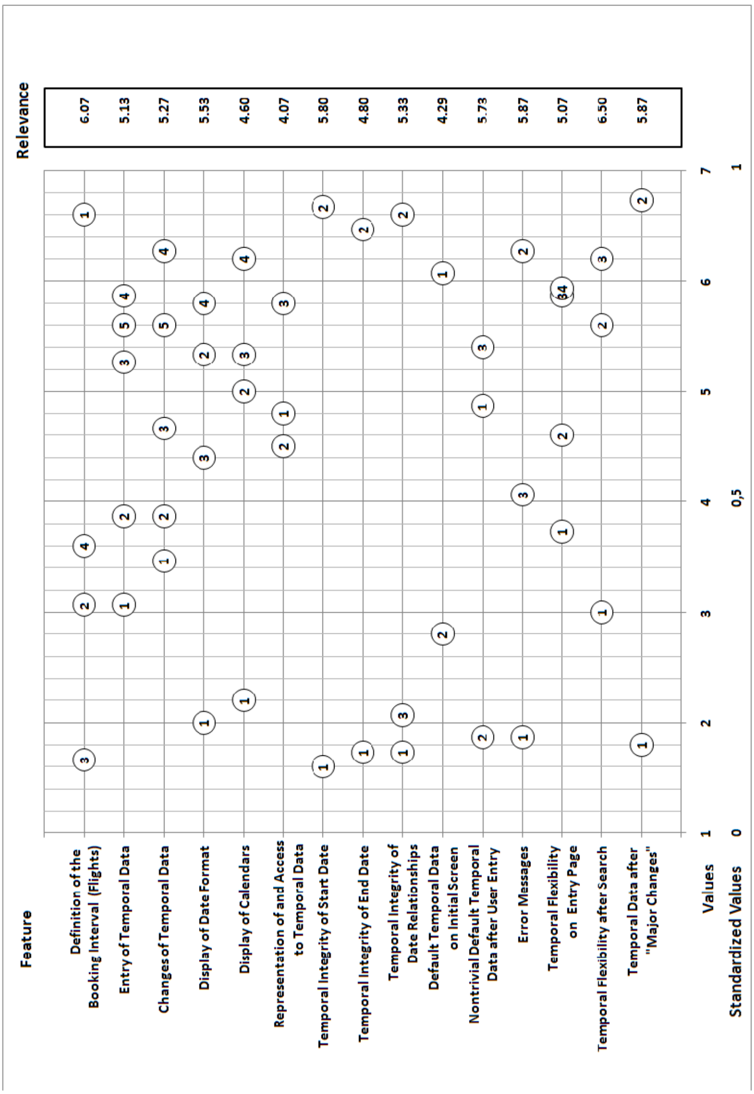


Fig. 1. Average Usability Evaluation by Usability Experts

We wanted to have all options j of feature i evaluated as a_{ij} on a 7 point scale ($1 \le a_{ij} \le 7$), in which a subset of values was verbally described as

$$a_{ij} = \begin{cases} 1 & \text{... totally unsuitable, annoying} \\ \ldots & \\ 4 & \text{... could be better, but acceptable} \\ \ldots & \\ 7 & \text{... perfect, no improvement potential} \end{cases}$$

The introduction to the questionnaire emphasized that if an expert cannot suggest a better option than those described in the Morphological Box, he should evaluate at least one of the options with 7. Unfortunately not all respondents followed this hint and we were not able to perform a second round of evaluations.

When we want to evaluate existing systems, not only the evaluation of the options is relevant but also an estimation of the relevance r_i of feature i. We therefore also asked each expert k how he estimates the relevance r_{ik} of feature i for users on a {1..7}-point scale.

4.2 Selected Results

Main results of the study are shown in Figure 1, which visualizes the average values obtained for each option. An option positioned in the right part of Figure 1 is evaluated as better usable than one which appears on the left side. Figure 1 also gives the average relevance factors the experts assigned to the features. In the following we describe some details of the evaluation results that seem worth to discuss.

Feature 1: Definition of the Booking Interval: In Table 1 we did not differentiate whether the booking should occur for a hotel or a flight. We assumed that there might be a difference, in particular with respect to the entry of the relevant booking interval. In Figure 2 we compare the average evaluations obtained for these two types of reservations. For hotel bookings defining the start day and the duration of the stay (Option 2 of Feature 1) is evaluated quite well (a_{12}=4.87) whereas for flights this option is regarded as inappropriate (a_{12}=3.07). We obtained no feedback that the difference between hotel and flight bookings may also be relevant for the other features. However, the differences shown in Figure 2 remind that the situational context of Human-Computer Interaction should always be kept in mind.

Rather surprisingly Option 4 for Feature 1 (entry allowed for all three temporal elements t_s, d, and t_e) was evaluated second-best for booking of flights. This evaluation may be based on the paradigm of recommending to give users as much freedom in communication with the system as possible. However, we are not sure whether the experts really were aware of the integrity problems resulting from allowing direct changes of all three temporal elements. If, for instance, the user entered t_s and d and the system computed t_e, how should the system react if the user changes t_s in a succeeding step: Should d or t_e be adjusted? In this context it is remarkable that the evaluation of Option 4 for Feature 1 showed the highest standard deviation of all evaluations (cf. Table 2).

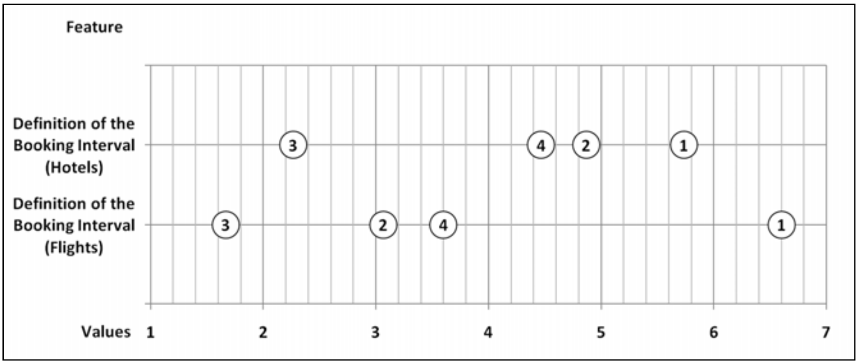


Fig. 2. Evaluations for Date Entry differ between Hotel and Flight Reservations.

Table 2. Options with Low and High Standard Deviation (Flight Reservations)

Feature	Option	Mean usability indicator value	Standard deviation of usability indicator value
15	2	6.73	0.59
9	2	6.60	0.74
7	2	6.67	0.82
…			
12	3	4.07	2.09
11	1	4.87	2.23
1	4	3.60	2.38

Feature 4: Display of Date Format: The experts favored a representation in which the months are abbreviated by letters (e.g., 05-Sep-2011), thus avoiding the problems related with pure numerical date representation (cf. [29, 30]). Such a representation is rarely used; it has the advantage that the intercultural differences of sequencing days and months [31] become irrelevant but may be difficult to interpret if the user has not sufficient skills in the language used on the Web site.

Feature 5: Display of Calendars: The experts preferred sites in which calendars are displayed without user action. Thus, they did not give too much weight to aspects of brimful pages and the potentially related information overload. Most websites do not present calendars immediately.

Feature 6: Representation of and Access to Temporal Data: The direct access to the month (Option 3 of Feature 6) is clearly evaluated best but also not realized too often. Direct access can be provided by dropdowns or by a clickable element of the calendar.

Features 7 to 9: *Temporal Integrity of Start Date, of End Date, and of Date Relationships:* With respect to the 3 temporal integrity features the experts favor representations that do not allow entering infeasible dates. If the user enters, e.g., a return date which lies before the outward date, the experts clearly refuse automatic adjustments without warnings.

Feature 12: Error Messages: In accordance with ISO 9241 but in contrast to the results presented in [32] the experts favored an immediate system reaction to entries that violate temporal integrity. However, if the user plans to correct both the start date and the end date of a reservation, an immediate error message may be inconvenient.

Features 13 and 14: *Travel Day Flexibility on Entry Page and/or after Search:* With respect to temporal flexibility on the entry page the experts prefer user-defined intervals (Options 4 and 3 of Feature 13) and give a high priority to showing several travel options around the originally selected dates after the search (Options 3 and 2 of Feature 14).

Feature 15: *Temporal Data after "Major Changes":* The loss of temporal data if the user changes, e.g., travel destinations is seen as inadmissible (Option 1 of Feature 15). If a system provides an explicit "go back" button, we used this button for back-tracking; in other systems we had to use the Browser's back button or even return to their homepages. Some of the losses of temporal data are based on "breaking the back button", which is regarded as a perpetual usability problem of Web systems [33].

Table 2 shows the standard deviations for the three options with the lowest and the three options with highest standard deviations. Least consensus between the experts exists with respect to the option of entering three temporal elements, of providing no nontrivial default values, and whether an error message should be delayed until a user action (such as Search). Most individual evaluations coincide quite well and their aggregates seem to be suitable as a basis for evaluating reservation systems.

When we look to the relevance indicators in the right column of Figure 1 we find that, on average, the experts evaluated all features at least with 4 on the {1..7}-point scale. The features considered most relevant are (in descending order)

- showing alternatives to the originally selected dates, preferably also with price information (Feature 14),
- the user-friendly definition of travel dates (Feature 1),
- immediate error messages (Feature 12),
- keeping temporal data in case of "major changes" (Feature 15), and
- (avoidance of) nontrivial default values (Feature 11).

5 An Empirically Based Evaluation Method

5.1 Defining Usability Indicators

For each option *j* of feature *i* we determined the average value

$$\bar{a}_{ij} = \sum_k a_{ijk} \,/\, K \qquad (1)$$

where a_{ijk} is the evaluation value assigned by expert *k* to option *j* for feature *i* and *K* is the number of evaluators. Furthermore, we computed the average value for relevance of feature *i* as

$$\bar{r}_i = \sum\nolimits_k r_{ik} \ / \ K \tag{2}$$

and the relative relevance of feature *i* as

$$\tilde{r}_i = \bar{r}_i \ / \ \sum\nolimits_i \bar{r}_i \tag{3}$$

The sum of the $\tilde{r}_i$ values adds up to one; we use these values as weights for the scores determined for Feature *i*.

In Figure 1 the options *j* for each feature *i* are positioned according to their $\bar{a}_{ij}$ values. If the profile lines for two systems differ and no point of the profile line for system X is left to any corresponding point of the profile line for system Y, system X dominates Y and is in all features evaluated as better or at least equally usable than system Y. However, when considering 15 features the dominance criterion will rarely hold.

We have no hints that the experts interpreted the scales used in the evaluation as nonlinear. Therefore we suggest computing a usability indicator M_s for system *s* which is determined as

$$M_s = \left[\sum\nolimits_i \left(\bar{a}_{ij(s)} * \tilde{r}_i\right) - 1\right] / \ 6 \tag{4}$$

where *ij(s)* symbolizes the option *j* realized by system *s* for feature *i*. The subtraction of 1 in formula (4) is necessary to transform values 1 to 7 to values 0 to 6; this transformation and the division by 6 lead to standardized usability indicator values between 0 and 1.

Even a system which realizes for each feature the best evaluated option would not receive an indicator value M_s=1 because this value can only be obtained if for each feature all experts prefer the same option and evaluate it with 7. Such consensus is highly unlikely. In our study the best evaluated options did not obtain average values of 7 but received values between 5.40 (Option 3 of Feature 11) and 6.73 (Option 2 of Feature 15).

An alternative scale results if we take the score of the best evaluations for each feature into account. Let h_i be the highest score obtained for feature *i* on the {1..7} scale. We can define an alternative usability indicator M_s' for system *s* as

$$M_s' = \left[\sum\nolimits_i \left(\bar{a}_{ij(s)} * \tilde{r}_i\right) - 1\right] / \left[\sum\nolimits_i (h_i * \tilde{r}_i) - 1\right] \tag{5}$$

which has the advantage that the maximum value 1 is achievable and, thus, may be more motivating in practical applications. The relative attractiveness of the systems is not influenced by this transformation because of the linear function

$$M_s' = \ \alpha \ M_s \tag{6}$$

When comparing selected systems with respect to their implemented options we used the indicator values M_s'.

5.2 Comparison of Selected Web-Based Reservation Systems

We selected the reservation systems of airlines to be analyzed with reference to several data sheets which define different classes of airlines (Table 3). The selection looks a bit arbitrary but is affected by the availability of data sheets. We were, e.g., interested whether very big airlines or highly rated airlines provide reservation systems that are more user-friendly with respect to handling temporal data than the systems of other companies. We had to adjust some of the lists; e.g., the list of the Top 10 Airlines in the USA and Canada (Class 3) mentions Skywest, which does not operate an own reservation system, and ExpressJet, which focuses on customized offerings for corporate clients and does not allow online booking. In the list of European Budget Airlines Scanderbeg seems no longer to exist and Eurofly was acquired by Meridiana fly. Pacific Blue is member of class 5 but its reservation system switches to Virgin Blue's system, which is also a member of class 5.

Many airlines belong to more than one class. For instance, Delta Air Lines is member of classes 1, 2, and 3; Emirates belongs to the classes 2, 4, and 5. Altogether we rated 60 airlines mentioned in at least one of the 6 classes defined in Table 3 for our statistics. However, additional evaluations were also conducted and some observations are mentioned later.

For each of the 60 airlines we analyzed its reservation system with regard to its treatment of temporal data. Whenever an English site was available, we evaluated this site. Because parallel sites of an airline sometimes do not always behave identical, this may introduce a bias against an airline which considers a site in its local language as its main site.

We found some cases which do not allow a unique marking in the Morphological Box. For instance, Delta's system shows a calendar as well after clicking into the date entry field as after clicking on a calendar icon (Feature 5). The Qantas system immediately adjusts the month if the user enters a return month prior to the departure month but does not immediately react if the return day entered is before the departure day of the same month (Feature 9). Temporal integrity may be checked after clicking calendar entries but not if the user changes the entry via the keyboard (Features 7 to 9). In such situations we split the assignments, in the case described with 0.7 for controlling temporal integrity for calendar entries and with 0.3 for neglecting it for keyboard entries. The effect of such splits can be seen, e.g., in Figure 3 where for Feature 7 the profile line for Delta Air Lines does not pass through a circle.

A dropdown usually impedes keyboard entry; but if a user of, e.g., Easyjet's system enters "J" in the month field via keyboard several times, the system switches between January, June, and July; if he enters, e.g., "2" in the day field, the display switches between 2, 20, 21, …, 29. With respect to displaying the date format (Feature 4) we found that Asiana Airlines displays the date (on its English site) in one field as YYYYMMDD, without any separator. Even if we would have anticipated such effects before designing the Morphological Box it would make no sense to define and evaluate rather implausible options. We even found differences in handling date relationships when we used different Browsers.

Table 4 shows the usability indicator values M_S' in descending order (as of March 31, 2011) for the 5 airlines which are ranked best, the 5 airlines evaluated least favorable, and for the airline which obtained an average evaluation with respect to treating temporal data.

Table 3. Classes of Airlines analyzed

# of class	Class	Link	# of airlines	Average M_s'
1	Scheduled Passengers Carried (as of 2010)	http://www.iata.org/ps/publications/Pages/wats-passenger-carried.aspx	10	0.776
2	Scheduled Passenger - Kilometres Flown (as of 2010)	http://www.iata.org/ps/publications/Pages/wats-passenger-km.aspx	9	0.811
3	Top 10 Airlines in the USA and Canada (Number of passengers) (as of 2010)	http://airtravel.about.com/od/basedinnorthamerica/tp/top10na.htm	8	0.800
4	The World's Top 10 airlines (2010 World Airline Awards)	http://www.worldairline awards.com/main/2010Awards.htm	10	0.746
5	Official World Airline Star Ranking (5 and 4 Star Airlines)	http://www.airlinequality.com/StarRanking/ranking.htm	36	0.732
6	Low Cost, Budget Airlines in Europe	http://airtravel.about.com/od/basedineurope/tp/lowcost budgetairlineseurope.htm	17	0.729

Table 4. Usability Indicators for selected Reservation Systems (as of March 31, 2011)

Reservation system of	Link	Member of classes	M_S'
Delta Air Lines	www.delta.com	1, 2, 3	0.934
American Airlines	www.americanairlines.ch/intl/ch/index_en.jsp	1, 2, 3	0.904
Virgin Atlantic	www.virgin-atlantic.com/en/eu/index.jsp	5	0.884
Lufthansa	www.lufthansa.com/de/en/Homepage	1, 2, 5	0.883
Jet Blue Airways	www.jetblue.com	3, 5	0.871
…..			
Finnair	www.finnair.com	5	0.738
…..			
Aer Lingus	www.aerlingus.com	6	0.660
Hongkong Airlines	www.hongkongairlines.com/eng	5	0.659
Hainan Airlines	global.hnair.com/en	5	0.604
Eva Air	www.evaair.com/html/b2c/english	5	0.576
China Southern Airlines	www.csair.com/en	1, 5	0.526

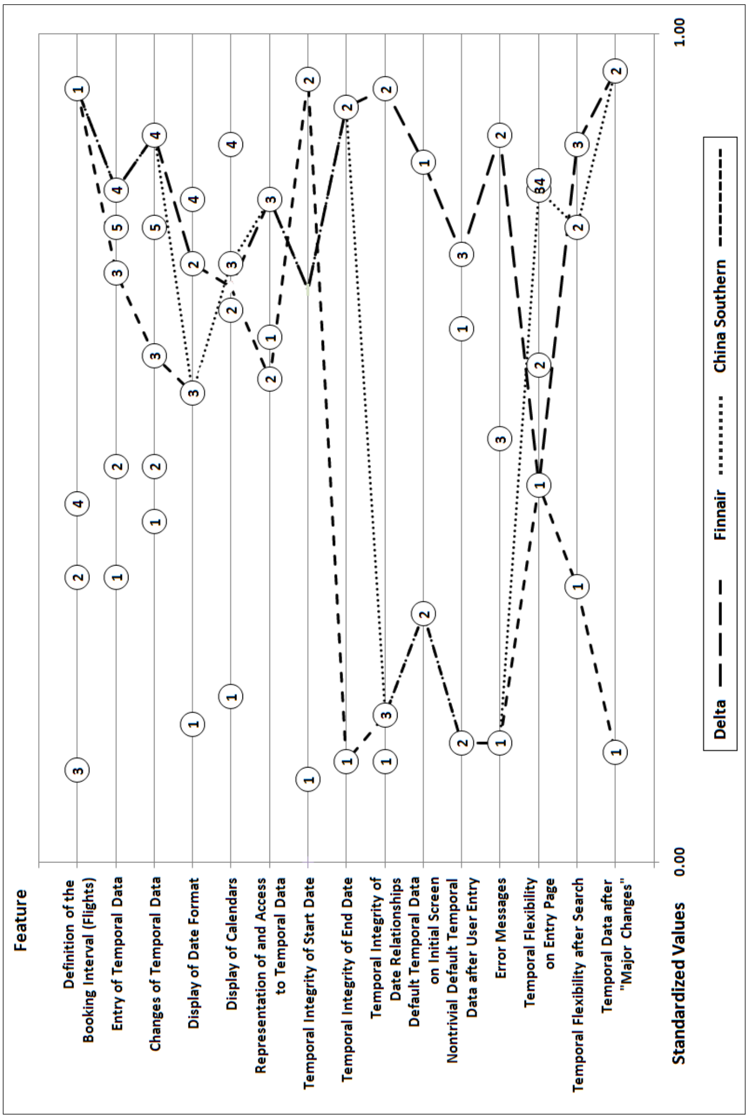


Fig. 3. Comparison of Options for treating Temporal Data in the Reservation Systems of Delta Air Lines, Finnair, and China Southern Airlines

Legend:

— · — · — · Identical profile segment for Delta's and Finnair's reservation systems

- · - · - · - Identical profile segment for Finnair's and China Southern's reservation systems

Figure 3 shows a detailed comparison of three reservation systems. We present the profile lines for the system which was evaluated best (Delta Air Lines), worst (China Southern Airlines), and for a system which represents average usability (Finnair). However, the "poorer" systems are not dominated by the "best" because of Features 5, 7, and 13.

According to our evaluation, Delta's system would achieve a usability indicator M_S' =1 if it would enhance the displays of the date and the calendar (Features 4 and 5), support temporal integrity also for keyboard entries (Feature 7), and offer travel day flexibility already on its entry page (Feature 13). Thus, even the system which was evaluated best did not realize the best evaluated options in 4 of 15 features.

The last column of Table 3 shows the average indicator values M_S' obtained by the airlines that are members of a certain class. Reservation systems of large carriers (Classes 1 to 3) obtain better average usability indicator values than those of less powerful carriers. Companies gaining awards (Classes 4 and 5) do not receive better evaluations than the large carriers. North American-based companies are evaluated better than those located in other regions.

It is worth mentioning that major differences between the usability values of systems exist even if the owners belong to the same group. For instance, Lufthansa owns Swiss and Austrian Airlines but the usability of their reservation systems differs. Table 5 summarizes the realized options and shows for which features we had to use split values. Only in 3 of 15 features all three systems realize the same option. None of the three airlines realized the best evaluated options for Features 5, 10, 11, and 13.

Table 5. Options realized by Airlines belonging to the same Group

Feature	**Options realized by**			
	Lufthansa	**Swiss**	**Austrian**	**"Best" option**
1	1	1	1	1
2	4	4	5	4
3	4	4	5	4
4	3	3	4	4
5	3	2	3	4
6	3	3	2	3
7	2	1 or 2	2	2
8	2	1 or 2	2	2
9	1 or 2	3	3	2
10	2	2	2	1
11	1	1 or 2	2	3
12	2 or 3	1 or 3	1	2
13	3	3	1	4
14	2	3	1	3
15	2	2	2	2
Usability indicator M_S'	0.883	0.774	0.687	

6 Summary and Conclusions

In this paper we enhanced previous research on usability aspects of handling temporal data by an empirical study in which usability experts were asked to evaluate the options defined in a Morphological Box. We discussed the results of the empirical study and used them to define usability indicators that allow evaluating the procedures implemented in different systems from a usability viewpoint. We computed usability indicator values for 60 reservation systems and developed profile lines for easy comparisons of systems. In all systems some improvement potential exists. The best results with respect to handling temporal data reach the systems of Delta Air Lines, American Airlines, Virgin Atlantic, Lufthansa, and Jet Blue Airways. We also showed that the systems provided by Lufthansa, Swiss, and Austrian Airlines differ remarkably in handling temporal data.

Even today information engineering concepts like temporal patterns or the reuse of existing services still seem to be widely neglected by the developers of Web-based reservation systems. The approach suggested in this paper may help system designers in selecting appropriate options for handling temporal data. It could also contribute to better conformance in properties where individual system design does not improve the market position of service providers but may prevent potential customers from booking if unfriendly interfaces annoy users.

References

1. Marcussen, C.H.: Trends in European Internet Distribution - of Travel and Tourism Services (2009), `http://www.crt.dk/uk/staff/chm/trends.htm`
2. Selvidge, P.: How Usable are the Airline Websites? Usability News 1 (1999), `http://www.surl.org/usabilitynews/11/Usability_airline.asp`
3. Crichton, E., Frew, A.J.: Usability of Information and Reservations Systems: Theory or Practice? In: Fesenmaier, D.R., Klein, S., Buhalis, D. (eds.) Information and Communication Technologies in Tourism 2000, pp. 408–417. Springer, Wien (2000)
4. Chariton, C., Choi, M.-H.: Enhancing Usability of Flight and Fare Search Functions for Air-line and Travel Web Sites. In: Proceedings of the International Conference on Information Technology: Coding and Computing (ITCC 2004), pp. 320–325. IEEE Computer Society, Los Alamitos (2004)
5. Malizia, A.: Adding flexibility to B2C booking systems using a virtual intermediate travel agent. In: Proceedings of the 2005 IEEE Symposium on Visual Languages and Human-Centric Computing, pp. 337–338. IEEE Computer Society, Los Alamitos (2005)
6. Zhou, Q., DeSantis, R.: Usability Issues in City Tourism Website Design: A Content Analysis. In: 2005 IEEE International Professional Communication Conference Proceedings, pp. 789–796. IEEE Press, Piscataway (2005)
7. Essawy, M.: Testing the Usability of Hotel Websites: The Springboard for Customer Relationship Building. Information Technology & Tourism 8, 47–70 (2006)
8. Lituchy, T.T., Barra, R.A.: International issues of the design and usage of websites for e-commerce: Hotel and airline examples. Journal of Engineering and Technology Management 25, 93–111 (2008)
9. Knolmayer, G.F., Helfenstein, L.E.: Usability Aspects of Ensuring Temporal Integrity Constraints: A Critical Appraisal. In: Modave, F., Dascalu, S., Moallem, A. (eds.) Computational Intelligence, Proceedings of the 5th International Conference on Human-Computer Interaction, pp. 271–278. ACTA Press (2010)

10. Grandi, F.: An Annotated Bibliography on Temporal and Evolution Aspects in the World Wide Web, TimeCenter Technical Report TR-75 (2003), http://timecenter.cs.aau.dk/TimeCenterPublications/TR-75.pdf
11. Grandi, F.: Introducing an Annotated Bibliography on Temporal and Evolution Aspects in the World Wide Web. SIGMOD Record 33, 84–86 (2004)
12. Jensen, C.S., Snodgrass, R.T. (eds.): Temporal Database Entries for the Springer Encyclopedia of Database Systems, TimeCenter Technical Report TR-90 (2008), http://timecenter.cs.aau.dk/TimeCenterPublications/TR-90.pdf
13. Liu, L., Özsu, T.M. (eds.): Encyclopedia of Database Systems. Springer, Berlin (2009)
14. Fabre, J., Howard, S.: Introduction to the Special Issue on Temporal Aspects of Usability. Interacting with Computers 11, 1–7 (1998)
15. W3C: Date formats (2007), http://www.w3.org/International/questions/qa-date-format
16. NHS: CAPS Development Lifecycle (2008), http://www.isb.nhs.uk/documents/cui/cuifullsubmission.pdf
17. Microsoft: Design Guidance, Date and Time Input, Version 3.0.0.0 (2010), http://www.mscui.net/DesignGuide/Pdfs/DesignGuidance-DateDisplay.pdf
18. Microsoft: Design Guidance, Date Display, Version 3.0.0.0 (2010), http://www.mscui.net/DesignGuide/Pdfs/DesignGuidance-DateandTimeInput.pdf
19. Bainbridge, A.: Hotel Date Entry: Design & Usability Report. Travel UCD (2002), http://www.tourcms.com/company/research/pdf/date_entry_hotel_july2002.pdf
20. Lennartz, S.: Online Calendars and Date Pickers.Smashing Magazine (2007), http://coding.smashingmagazine.com/2007/10/23/online-calendars-and-date-pickers/
21. Fowler, M.: Temporal Patterns (2005), http://www.martinfowler.com/eaaDev/timeNarrative.html
22. Zwicky, F.: Morphological Astronomy. The Observatory 68, 121–143 (1948)
23. Zwicky, F.: Discovery, Invention, Research Through the Morphological Approach. Macmillan, Toronto (1969)
24. Knolmayer, G., Borean, A.: A Morphological Box for Handling Temporal Data in B2C Systems. In: Cellary, W., Estevez, E. (eds.) Proceedings of the 10th IFIP Conference on e-Business, e-Services, and e-Society, Software Services for e-World, pp. 215–225. Springer, Berlin (2010)
25. Rosson, M.B., Carroll, J.M.: Usability Engineering. Morgan Kaufmann, San Francisco (2002)
26. Virzi, R.A.: Refining the test phase of usability evaluation: How many subjects is enough? Human Factors 34, 457–468 (1992)
27. Nielsen, J.: How to Conduct a Heuristic Evaluation (no date), http://www.useit.com/papers/heuristic/heuristic_evaluation.html
28. Cockton, G., Woolrych, A., Lavery, D.: Inspection-Based Evaluations. In: Sears, A., Jacko, J.A. (eds.) The Human-Computer Interaction Handbook, 2nd edn., pp. 1171–1208. CRC Press, New York (2008)
29. Nielsen, J.: Designing Web Usability. New Riders, Indianapolis (2000)
30. Nielsen, J., Tahir, M.: Homepage Usability. In: 50 Websites Deconstructed. New Riders, Indianapolis (2002)
31. Hanna, J.: Date & Time Formats on the Web (no date), http://www.hackcraft.net/web/datetime/
32. Bargas-Avila, J., Oberholzer, G.: Online form validation: Don't show errors right away. In: Rauterberg, M., Menozzi, M., Wesson, J. (eds.) Human-Computer Interaction, INTERACT 2003, pp. 848–851. IOS Press, Amsterdam (2003)
33. Nielsen, J., Loranger, H.: Prioritizing Web Usability. New Riders, Berkeley (2006)

A Tool Support for Web Applications Adaptation Using Navigation History

Sergio Firmenich[1], Marco Winckler[2], and Gustavo Rossi[1]

[1]LIFIA, Facultad de Informática, Universidad Nacional de La Plata and Conicet Argentina
{sergio.firmenich,gustavo}@lifia.info.unlp.edu.ar
[2] ICS-IRIT, Université Paul Sabatier, France
winckler@irit.fr

Abstract. Currently the Web is a platform for performing complex tasks which involve dealing with different Web applications. However users still have to face these tasks in a handcrafted way. In this paper we present a novel approach that combines concern-sensitive adaptation and navigation history to improve the user experience while performing a task. We have developed some simple though powerful tools for applying this approach to some typical tasks such as trip planning and house rental. We illustrate the paper with a simple though realistic case study and compare our work with others in the same field.

1 Introduction

One of the interesting facets of Web evolution is the end-users interaction with Web content [2]. At first, users could only browse through contents provided by the web site. Later, users could actively contribute with content by using tools (e.g. wikis) embedded into the web site. Recent technologies provide users with tools for changing the way Web content is displayed. For example, visual Mashups [7, 17], support the integration of content hosted by diverse web sites and Greasemonkey scripts [10] allow users to change third part web applications by adding content and/or controls (e.g. to highlight search results in Amazon.com which refer to Kindle).

These tools follow the concept of *Web augmentation* [4] by extending what users can do with Web content. However, they provided limited support to tasks requiring navigation of many Web sites. For example, a user planning a holiday trip to Paris might ultimately visit several web sites such as *expedia.com* for flights, *booking.com* for hotels, *wikipedia.org* for general information about the city and *parisinfo.fr* for points of interest, current events or expositions in Paris... From the users' point of view, the navigation of all these web sites is part of the same task. The existing augmentation techniques are of little help in this case. GreaseMonkey scripts can adapt the content on a specific Web site but it will require much effort to make it generic enough to integrate information provided by different applications. Mashup for *expedia.com* will not necessarily integrate information from users' preferred web sites (e.g. *airfrance.fr*, *venere.com...*). If Web sites provide public APIs, Mashups can be extended but it does not prevent users to learn how to do it beforehand. Quite often, users' tasks are associated with opportunistic navigation on different Web sites, which is difficult to predict [14]. In this context, effective Web augmentation should

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 340–348, 2011.

overcome to main barrier: i) to take into account the different web sites visited by users; and ii) to adapt target web sites accordingly to unpredictable user needs.

This paper investigates the use of a tool support for creating flexible, light-weight and effective adaptations to support users' tasks during the navigation of diverse Web applications. Our goal is to support users' tasks by keeping his actual concern (and related data) persistent through applications. For example, allow the reuse of dates provided on *expedia.com* for booking a flight to search hotels at *booking.com.* Another example, allow the inclusion of new links letting users to navigate from *parisinfo.fr* to related articles at *wikipedia.com* whenever he needs further explanation about a topic. Hereafter we present the tools we have developed to solve this kind of problem. Section 2 shows a view at glance of our approach for Client-Side Adaptation (CSA) of Web applications. Section 3 presents the tools we have developed. Section 4 presents how we have validated our approach with end-users. Section 5 discusses related work and section 6 presents conclusions and future work.

2 The Underlying Approach

Our approach is based on concept of concern-sensitive navigation. We say that a Web application (or specifically a Web page) is concern-sensitive (CS) when its contents, operations and outgoing navigation links can change (or adapt) to follow the actual situation (concern) in which it is accessed [9]. Concern-sensitive navigation is different from context-aware navigation, where other contextual parameters (location, time, preferences) are considered. The main difference is that concern-sensitive is driven by a specific user goal, quite often volatile and difficult to generalize.

Figure 1 shows an example of concern-sensitive navigation across two applications: Google Maps (as the source of navigation) and Wikipedia (as the target). The left-side of Figure 1 displays Wikipedia links in the map of Paris; once selected, these links trigger the exhibition at the right-side of Figure 1 the corresponding map and a set of links to those Wikipedia articles in the surroundings of the current one.

Fig. 1. Inter-application CSN between Google Maps and Wikipedia

We assume that concern-sensitive navigation simplifies the user's tasks by providing him sensitive information or options according to his current needs. For that purpose, the adaptation of a page *P* requires that: (a) the actual user's navigation concern (i.e. pages previously navigated, e.g. Google maps), (b) the set of relevant information from previously visited pages that are needed for adaptation (e.g. the current map), and (c) the capacity for enriching *P* with contents or links related with (a) and (b) by intervening in *P*'s DOM.

3 Tool Support

The tool presented hereafter was implemented as a Firefox plug-in that provides components called *augmenters*. In our approach, users need to run adaptations (using *augmenters*) on different Web sites they are visiting to perform their tasks. This implies that users must to collect information during the Web sites navigation. A set of tools called *DataCollectors* work as a memory for user data. The other *augmenters* will then use information stored by *DataCollectors* to perform the adaptations.

3.1 Data Collector and Other Augmenters

Two types of *DataCollectors* have been implemented: *Untyped Pocket* which implements a *copy* & *paste* behaviour for simple text (e.g. "*Paris*"); and *Typed Pocket* which allows users to label data (e.g. "*Paris*" is "*City*"). In both cases, selected information is placed into a temporary memory called *Pocket*. Data collection is supported via the contextual menu options "*put it into pocket"* (i.e. *Typed Pocket*) and *"put it into volatile pocket"* (i.e. *Untyped Pocket*). Figure 2.a illustrates the collection of a piece of information (i.e. *"Place de la Concorde"*) using the option *"put it into pocket"* which has been labelled "*PointOfInterest"*. Collected information become available into the Pocket as shown by Figure 2.b (yellow box at the upper-left side).

Fig. 2.a. Information collection from Wikipedia using a *DataCollector*

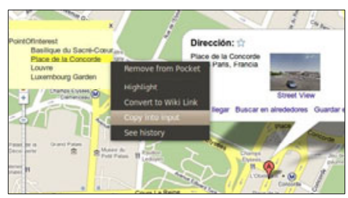


Fig. 2.b. Resulting adaptation for the *Pocket* memory

The *Pocket* is more than just a simple post-it as the information it stores can be used by *augmenters* to adapt Web sites. Currently available *augmenters* include:

- *Highlight*: it colors the occurrences of the data received by parameter.
- *CopyIntoInput*: it pastes the value received as parameter into an input form field. Once executed, *CoopyIntoInput* adds a listener to the click event which is removed after the first time that the target is an input.
- *WikiLinkConvertion*: it creates links to *wikipedia.com* pages using as input any occurrences values received as parameter. For example if the parameters is "*Paris*" then the link would be to the Wikipedia article about Paris.

An example of how the *augmenters* work is provided by Figure 2. At Figure 2.a the user has used a *DataCollector* augmenter to capture information at the web site *wikipedia.org*. When the user opens the *Google Maps* web site (Figure 2.b) the information collected in previous Web site (i.e. wikipedia.org) is available at the *Pocket* (the yellow box at left in Figure 2.b). Now, the contextual menu at Figure 2.b offer new augmenters based on the previous collected information, *CopyIntoInput*, *Highlight* and *WikilinkConvertion*. However which augmenters are available depend on the current site because augmenters can be generic enough to be applied to any page (e.g. *highlight*). These augmenters are illustrated by the scenarios below.

3.2 Scenario for Performing Adaptations

The scenario presented in this section aims at fulfilling users' needs described at section 1. While booking flights to Paris, the user collects data (cf. Figure 3.a) which will help him in the next steps to find a hotel. Relevant information is labelled by the user as *departDate*, *arriveDate* and *destination*. Figure 3.b shows how the form field *destination* is filled in with the information previously collected. This scenario is executed once the user reaches the page *booking.com* (either by following a link or entering a new URL). Notice that the scenario can be instantiated because the information needed is available into de *Pocket*. So far, automatic form filling can only be done for a particular Web application (in our case for *booking.com*) but this feature can be extended using tools like Carbon [1]. This use of concern-sensitive information improves the user experience by allowing him to "transport" critical data among Web applications and use these data to adapt them.

Fig. 3.a. Information extraction from *expedia.com*

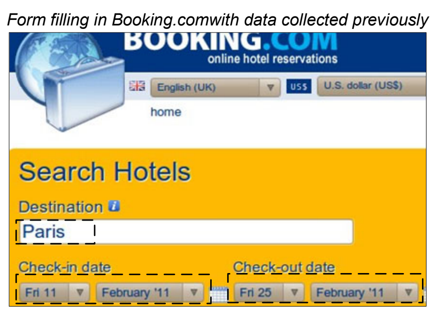


Fig. 3.b. Form filling in *booking.com* with information collected in previous web sites

Figure 4 exemplifies the use of the augmenter *WikiLinkConvertion*. In this scenario, we assume the user has previously visited the web site *Wikipedia.com* and selected his *PointOfInterest*. Then the user opens the web site *Parisinfo.com*. It is worthy noting that the *Pocket* features all information previously collected at the web site *Wikipedia.com.* Now, the user right clicks over *PointOfInterest*, a contextual menu offer the option "*Convert to Wiki Link*" that, once selected by the user, will run the augmenter *WikiLinkConvertion,* thus creating new links on the current page of the web site *ParisInfo.com* allowing the navigation to the web site *Wikipedia.com* to the corresponding page associate to the information *PointOfInterest*. The augmenter *Highlight* works in a similar way but it changes the colour of the text instead of created links. Due to space reasons, the augmenter *Highlight* is not illustrated here.

Fig. 4. Text plain converted into link to add personal navigation

4 Evaluation of the Approach

To validate our approach and actual usage of the tools, we have conducted a usability study with end-users. The goal of this evaluation was to investigate if CSA is usable for solving common tasks whilst navigating web. The adaptations investigated in this study explored the following augmenters: *Highlight* for changing color of important information, *WikiLinkConvertion* for creating new links to Wikipedia, *DataCollector* for recording information for later usage, and *CopyIntoInput* for automating filling in forms.

The study was run with 11 participants (6 males and 5 females, aged from 23 to 46 years old). All participants were experienced Web users (i.e. > 5 years using the web) that browse the web as part of their daily activities (in average 4,1 hours of navigation on the web per day, SD=2,4 h). We have focused on experienced users because we assume that they are more likely to formulate special needs for adapting Web pages than novices with the Web. Participants were asked to fill out a pre-questionnaire, following they were introduced to the system (i.e. 2-5 minutes training) and asked to conduct five tasks at their workplace, followed by a final interview and a System Usability Scale questionnaire (i.e. SUS, [5]). The SUS has been as a complement to user observation as it is widely used in comparative usability assessments in the industry.

The five user's tasks concern a trip planning to Paris for visiting the art exhibition "*De Stijl et Mondrian*". The initial setup was a Web page advertising that art exhibition. The tasks were: 1) to collect required data for planning the trip including dates, keywords and locations; 2) to book a hotel in Paris near the exposition for the week-end of February 18th 2011; 3) to select a hotel in the neighborhood of "*Les Marais*"; 4) to record information about the hotel; 5) to create a relationship between the actual Web of the exhibition and the Web site *wikipedia.com*.

Usability was measured in terms of time to accomplish tasks, number of tasks performed successfully, and user satisfaction (via a questionnaire). Users were also asked to rate every task from 1 to 5 (from very easy to very difficult).

All participants used the tools presented during the training period to perform the tasks. Users completed the tasks in approximately 37 minutes (SD=9 minutes). The results show that, generally, participants appreciate the concept of CSA and the tool support. In the pre-questionnaire, when asked if they would like to modify the web pages they visit, 2 of 11 participants said no because "it could be very time consuming". Notwithstanding, all participants said that our tools for client-side adaptation are useful and that they are willing to use it in the future. Adaption across different web site was described as "natural" by 7 participants and a "real need" by 5 of them. The tool *DataCollector* was the most successful applied by all participants; it was considered the very useful and a "good substitute for post-its". However, success rate varied according the augmenter employed: *CopyIntoInput* was considered very easy to use by participants and employed successfully by 10 of them (90,9%). The augmenter *Highlight* (72% of success rate, 8 participants) was considered easy to use but 5 users blamed it because they only can be applied to the exact word previously selected and users cannot choose the color and/or the policy used to highlight different pieces of information. Participants were very impressed by the augmenter allowing links to Wikipedia from concepts, i.e. *WikiLinkConvertion*; despite the fact it was considered extremely useful, the success rate with this augmenter was the lowest in the study, i.e. 18%, due to two main issues: the fact that links can only be created from typed information; lack of visual feedback (i.e. an icon) indicating where that action was possible. Nine participants (81,8%) mentioned that using the augmenters improve their performance with tasks, one user said it could be faster without the augmenters and the other one didn't see any difference. This user perception has been confirmed by the time recorded during task execution using augmenters *WikiLinkConvertion* and *CopyIntoInput*.

This study also revealed some usability problems that motivate further development in the tool. For example, users requested to have a visual indicator allowing them to distinguish where augmenters have been applied (ex. links on the web site x links created with the augmenter *WikiLinkConvertion*). Users intuitively tried to activate some of the augmenters using *Drag & Drop* which is an indicator for further research of more natural interaction with *augmenters*. The most frequent suggestions for new augmenters include "automatic filling forms", "create links to other web sites than Wikipedia", and "automatic highlight at the web page of information previously collected". This positive analysis is confirmed by a SUS score of 84,9 points (SD = 5,5), which is a good indicator of general usability of the system.

5 Related Work

The field of Web applications adaptation is broad; therefore, for the sake of conciseness we will concentrate on those research works which are close to our intent. The interested reader can find more material on the general subject in [6]. As stated in the introduction we can identify two coarse-grained approaches for end-user development in Web applications: i) mashing up contents or services in a new application and ii) adapting the augmented application, generally by running adaptation scripts in the client side.

Mashups are an interesting alternative for final users to combine existing resources and services in a new specialized application. Visual and intuitive tools such as [16] simplify the development of these applications. Since most Web applications do not provide Web services to access their functionality or information, [11] proposes a novel approach to integrate contents of third party applications by describing and extracting these contents at the client side and to use these contents later by generating virtual Web services that allow accessing them.

The second alternative to build support for their tasks is Web augmentation [4], where applications are adapted instead of "integrated" in a new one. This approach, as indicated in [2] is very popular since it is an excellent vehicle for crowdsourcing. Many popular Web applications such as Gmail have incorporated some of these user-programmed adaptations into their applications like the mail delete button (See http://userscripts.org/scripts/show/1345). GreaseMonkey [10] is the most popular tool for Web augmentation, and its scripts are written in JavaScript. The problem with these scripts is their dependence on the Document Object Model (DOM) used to organize the Web page; if the DOM changes the script can stop working. In [8] the authors propose a way to make GreaseMonkey scripts more robust, by using a conceptual layer (provided by the Web application developer) over the DOM.

While we share the philosophy behind these works, we believe that it is necessary to go a step further in the kind of supported adaptations. In [9] we showed how to use the actual user concern (expressed in his navigational history) as an additional parameter to adapt the target application. By using the scripting interface we managed to make the process more modular and by defining adaptations for application families (e.g. social networks) we improved the reuse of adaptation scripts. In the following sections we show how to broaden the approach allowing end users to select which concrete information can be used to perform the adaptation, therefore improving the support for his task and providing support for building more complex adaptations. Some tasks are repeated several times and then users make the same process each time. This problem has been tackled in [3,13] with the CoScripter tool. CoScripter is a Firefox Plug-in which allows users to record his interactions with a Web Site, and then, they can repeat the process automatically later. The approach is a bit flexible, for example, the whole process can be repeated with other information in form inputs that those used in the original recorded execution, but always using the same fixed Web sites. In this way, CoScripter is not useful when users need to change slightly the process, for example by changing which is the target Web application. However, both CoScripter's goals and our approach's goals are different, because with CoScripter Web applications are not adapted (not further that fill forms with values) and volatile requirements are not contemplated.

Other near work is [15], which is other Firefox plug-in addressed to improve user experience by empowering his browser with commands with different goals. With MozillaUbiquity users execute commands (developed by themselves) for specific operation, for example to publish some text from the current Web page in a social network. Anyway, these commands are executed under user demand, and adaptations are not made automatically. Although MozillaUbiquity makes short the distance between two distinct Web Applications, to move information from one of them to another is not fully exploited.

6 Conclusions and Future Work

In this work we have presented a novel approach of CSA driven by the integration of information through the navigation of several web applications. The underlying idea is to support concern-sensitive adaptations on Client-Side in order to improve users experience while they are doing tasks in many web sites. The tools presented in this paper are based on a framework (not presented here) that allows the development of augmenters beyond those presented. The present study allows us to investigate new strategies of *Web augmentation* with end users. A user testing experiment was performed to demonstrate the feasibility of the strategy of CSA. Despite some usability problems found with the actual tools, the preliminary results show that approach is very promising and can indeed help users to solve complex tasks that require information exchange between different web sites. We observe an increasing interest in the development of tools that can make users more active with respect the way they access content provided by Web applications. Notwithstanding, there is a few empirical studies that investigate the user experience of user driven CSA, in particular when adapting third-party web sites. As far as the adaptation across different web sites is a concern, we haven't found in the literature any other tool allowing users to freely adapt web pages accordingly to previously navigation of web pages. The results presented here remains preliminary but it provides many insights for discussions, including: users' needs for performing complex tasks among different web sites, development of new strategies of end-user programming of the web, impact of user-driven adaptation of web site. Our next steps include the investigation of CSA beyond a single user session, for example, when navigation of different web sites occurs in a long period of time.

References

1 Araújo, S., Gao, Q., Leonardi, E., Houben, G.: Carbon: Domain-Independent Automatic Web Form Filling. In: Benatallah, B., Casati, F., Kappel, G., Rossi, G. (eds.) ICWE 2010. LNCS, vol. 6189, pp. 292–306. Springer, Heidelberg (2010)

2 Arellano, C., Díaz, O., Iturrioz, J.: Script Programmers as Value Co-creators. In: Daniel, F., Facca, F.M. (eds.) ICWE 2010. LNCS, vol. 6385, pp. 417–420. Springer, Heidelberg (2010)

3 Bogart, C., Burnett, M., Cypher, A., Scaffidi, C.: End-user programming in the wild: A field study of CoScripter scripts. In: Proceeding of EEE Symposium on Visual Languages and Human-Centric Computing, Germany, pp. 39–46 (2008)

4 Bouvin, N.O.: Unifying Strategies for Web Augmentation. In: Proc. of the 10th ACM Conference on Hypertext and Hypermedia (1999)

5 Brooke, J.: SUS: a 'quick and dirty' usability scale. In: Usability Evaluation in Industry, Taylor and Francis, London (1996)

6 Brusilovsky, P.: Adaptive Navigation Support. In: The Adaptive Web: Methods and Strategies of Web Personalization, pp. 263–290. Springer, Heidelberg (2007)

7 Daniel, F., Casati, F., Soi, S., Fox, J., Zancarli, D., Shan, M.: Hosted Universal Integration on the Web: The mashArt Platform. In: Proceeding of ICSOC/ServiceWave, Stockholm, pp. 647–648 (2009)

8 Diaz, O., Arellano, C., Iturrioz, J.: Layman tuning of websites: facing change resilience. In: Proc. of WWW 2008 Conference, Beijing, pp. 127–1128 (2008)
9 Firmenich, S., Rossi, G., Urbieta, M., Gordillo, S., Challiol, C., Nanard, J., Nanard, M., Araujo, J.: Engineering Concern-Sensitive Navigation Structures. Concepts, tools and examples. In: JWE 2010, pp. 157–185 (2010)
10 Greasemonkey, http://www.greasespot.net/ (last visit on June 7, 2011)
11 Han, H., Tokuda, T.: A Method for Integration of Web Applications Based on Information Extraction. In: Proceeding of ICWE, New York. Springer, Heidelberg (2008)
12 Ikeda, S., Nagamine, T., Kamada, T.: Application framework with demand-driven mashup for selective browsing. In: Proc. of the 10th Int. Conf. on Information Integration and Web-based Applications & Services (iiWAS 2008), pp. 33–40. ACM, New York (2008)
13 Leshed, G., Haber, E., Matthews, T., Lau, T.: CoScripter: automating & sharing how-to knowledge in the enterprise. In: Proc. of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems, Italy, pp. 1719–1728 (2008)
14 Miller, C.S., Remington, R.W.: Modeling an Opportunistic Strategy for Information Navigation. In: 23th Conf. of the Cognitive Science Society (2001)
15 MozillaUbiquity, http://mozillalabs.com/ubiquity/ (last visit on June 7, 2011)
16 Yu, J., Benatallah, B., Casati, F., Daniel, F.: Understanding Mashup Development. IEEE Internet Computing 12, 44–52 (2008)
17 Wong, J., Hong, J.I.: Making Mashups wit Marmite: Towards End-User Programming for the Web. ACM, City (2007)

Web Usability Probe: A Tool for Supporting Remote Usability Evaluation of Web Sites

Tonio Carta[1], Fabio Paternò[1], and Vagner Figuerêdo de Santana[1,2]

[1] CNR-ISTI, HIIS Laboratory, Via Moruzzi 1, 56124 Pisa, Italy
[2] Institute of Computing - University of Campinas, Albert Einstein Av., 1251, Campinas, São Paulo, Brazil
{Tonio.Carta,Fabio.Paterno}@isti.cnr.it, vsantana@ic.unicamp.br

Abstract. Usability evaluation of Web sites is still a difficult and time-consuming task, often performed manually. This paper presents a tool that supports remote usability evaluation of Web sites. The tool considers client-side data on user interactions and JavaScript events. In addition, it allows the definition of custom events, giving evaluators the flexibility to add specific events to be detected and considered in the evaluation. The tool supports evaluation of any Web site by exploiting a proxy-based architecture and enables the evaluator to perform a comparison between actual user behavior and an optimal sequence of actions.

Keywords: Tools for Usability Evaluation, Remote evaluation, Log analysis.

1 Introduction

Although usability has long been addressed and discussed, when people navigate the Web they often encounter a number of usability issues. This is also due to the fact that Web surfers often decide on the spur of the moment what to do and whether to continue to navigate in a Web site. Usability evaluation is thus an important phase in the deployment of Web applications. For this purpose automatic tools are very useful to gather larger amount of usability data and support their analysis.

Remote evaluation [2] implies that users and evaluators are separated in time and/or space. This is important in order to analyse users in their daily environments and decreases the costs of the evaluation without requiring the use of specific laboratories and asking the users to move. In addition, tools for remote Web usability evaluation should be sufficiently general so that they can be used to analyze user behavior even when using various browsers or applications developed using different toolkits. This work presents Web Usability Probe (WUP), a tool that follows a proxy-based architecture, performs remote evaluation, and considers client-side logs as data source. We prefer logging on the client-side in order to be able to capture any user-generated events, which can provide useful hints regarding possible usability problems. Moreover, WUP allows the usability experts to analyze through some graphical representations of the logged data how users interacted with the user interface (UI).

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 349–357, 2011.

Ivory and Hearst [4] provided a good discussion of tools for usability evaluation according to a taxonomy based on four dimensions: method class (the type of evaluation); method type (how the evaluation is conducted); automation type (the evaluation aspect that is automated); and effort level (the type of method required to execute the method). According to this classification, the WUP solution for usability testing involves: capturing logs generated at client-side, supporting automatic analysis and a number of visualizations to ease the identification of the usability issues, and only requiring that users perform some predefined tasks specified by the evaluators.

Google Analytics [1] has the potential to be configured to capture custom events at client-side and it offers a number of statistics information and reports, but it is rather limited in terms of number of events that it captures for each session. Model-based approaches have been used to support usability evaluation. An example was WebRemUsine [6], which was a tool for remote usability evaluation of Web applications through browser logs and task models. Propp and Frobrig [7] have used task models for supporting usability evaluation of a different type of application: cooperative behavior of people interacting in smart environments. A different use of models is in [5], in which the authors discuss how task models can enhance visualization of the usability test log. In our case we do not require the effort of developing models to apply our tool. We only require that the designer provides an example of optimal use associated with each of the relevant tasks. WUP will then compare the logs with the actual use with the optimal log in order to identify deviations, which may indicate potential usability problems.

2 The Approach Proposed

The approach proposed is based on an intermediate proxy server whose purpose is to annotate the accessed Web pages in order to include JavaScripts that will carry out the logging of the actual user behavior (see Figure 1). WUP does not require use of plug-in installation or specific client configuration. The scripts used by the tool are stored in the proxy server and thus there is no security conflict when the page is accessed from the client since the page appears as coming from the proxy server.

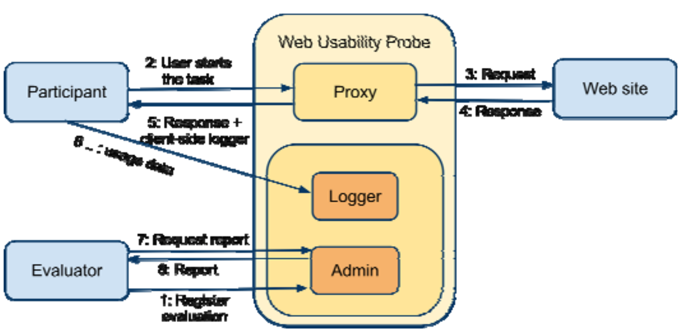


Fig. 1. Overview of the approach proposed in the Web Usability Probe

The WUP's server also acts as a usability server in which the evaluator can enter information useful to guide the usability test, such as the list of tasks to perform, what events to log, etc. Once users have carried out their test tasks, the logs can be accessed by the evaluators who can customize their representations by selecting what actions to be represented in the reports in order to ease the identification of usability issues.

When a new user test is to be started, the evaluator can access the tool to provide users with the information required to carry out the test (task descriptions, etc.). The evaluator then plays the role of user by performing the tasks in an optimal way, without errors (which are actions useless for the current task). The optimal task is defined during the planning phase of user test and the evaluator has to perform it according to tasks' descriptions, considering natural usage, and without useless actions. Users only need to start using the Web site to be evaluated through the proper link available at the proxy. Then, when users start the user test they are informed about the task to accomplish. When they complete it, they have to indicate this by accessing the dialogue box automatically generated with the controls related to the task. Next, a new task to perform will be indicated until the end of the test is reached.

The scripts that are injected by the server in order to log usage data and always redirect navigation through the proxy are implemented in JavaScript and use jQuery[1]. All the logged data are sent asynchronously to the server while the users move from one page to another.

This approach satisfies a number of requirements about evaluation tools [8]: it works in different configurations of hardware and software; it does not depend on specific configurations; it does not impact on the Web site usage; and it does not interfere with the Web page.

The development of a proxy-based tool considering client-side data encounters different challenges regarding the identification of the elements that users are interacting with, how to manage element identification when the page is changed dynamically, how to manage data logging when users are going from one page to another, among others. The following are some of the solutions we adopted in order to deal with these issues.

When logging data at the client-side, identification of the target element is an issue if the target element of a certain event does not have a name or the *id* attribute. In this case, two main approaches can be followed: either ignore events involving unidentified elements or assign an *id* attribute according their position in the Web page structure, i.e., DOM (Document Object Model) Tree. The first approach does not allow the evaluator to know exactly the elements referenced by the trigged events if any usability problem is identified, which complicates its correction. The second approach, on the other hand, can cause some overhead. The approach used in the WUP generates *ids* according to the XPath[2] syntax, and thereby allowing the *id* attribute to be computer and human readable. This provides the possibility of mapping back an *id* to the Web page element being analysed.

Another challenge that a client-side logging tool should address is how to inform evaluators that only a certain part of the page has been changed dynamically. This can

[1] `http://jquery.com`
[2] `http://www.w3.org/TR/xpath/`

occur, for instance, via AJAX, which can result in new UI elements. In this case the event is triggered when the DOM Tree is changed, and it can be reported in the timeline, allowing the evaluators to verify where and when it occurred, helping evaluators to deal with the issue of identifying UI elements users interact with.

3 Setting Up a Remote Usability Test

The evaluators can create the settings for a remote usability test at any time. In the administrator part of the usability tool there is an item for each test indicating the name, the description, the evaluator who created it, and the number of tasks that should be performed. The tool provides dynamically the indication of the number of sessions that have been logged for that test.

Associated with each test there is an editable list of tasks (an example in Figure 2). For each task there is the indication of the name, the description, the indication of the URL where the task should be started, whether it is skipable, and if its performance depends on some other task. A dependency means that the other task should be performed first than the current one.

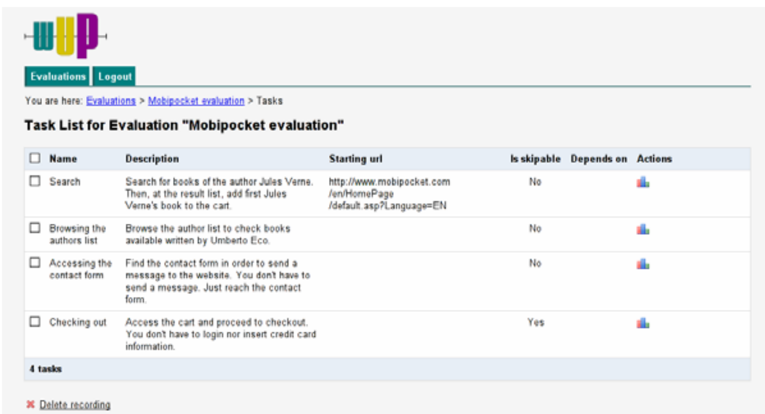


Fig. 2. An example of task list for a user test

These features aim to allow a remote participant to freely perform tasks. The name and description are the texts that users will see through automatically generated dialogue boxes. The starting URL provides flexibility in the evaluation setup, since allows evaluators to define that certain tasks have to start in different parts of the evaluated Web site, or even start in another Web site, allowing evaluations to perform comparison among Web sites.

Finally, the dependency feature among tasks is provided in order to make possible to define evaluations where one tasks is mandatory (e.g., login) in order to perform others (e.g., creation of content in a login protected Web site).

The vocabulary of events supported by WUP is one of our contributions, since it captures any standard event[3], jQuery Events[4], touch, gestures, and accelerometer events present at the Safari API[5]. The set of events observed by WUP is shown by grouping them by their type (defined according to the device that generate them), for instance: accelerometer, keyboard, mouse, touch. We also consider form-related events (e.g., change, select, and submit), system related events, and customizable events. The evaluator can define custom events, which can be various types of composition of basic events in terms of their ordering or standard events on specific parameters (e.g. a pageview event is triggered when a Web page is shown to the user), and it is possible to associate them with specific events names that can then be visualized in the reports. Moreover, the pageview custom event allows WUP to log information commonly present at server logs, since this custom event counts with URL path and its parameters.

4 Analysis of a Remote Usability Test

Once some users have actually performed the user test, the evaluator can access graphical representations of such logs. Hilbert and Redmiles [3] indicated that timelines can be useful for this purpose. We follow this approach: there is one timeline for each task performed by a user (see Figure 3). The first timeline is dedicated to the optimal log. The graphical representation is interactive and allows the designers to line up logs according to some important event under the 'lock scroll' UI control. In this way the evaluator can compare the optimal behavior with those of the actual users in order to see whether there was some particularly long interaction, due, for example, to the difficulty to understand how to proceed, or some errors that indicate some usability issue. In addition, the zoom level of timelines can be interactively set in order to identify the more adequate representation scale of the timeline.

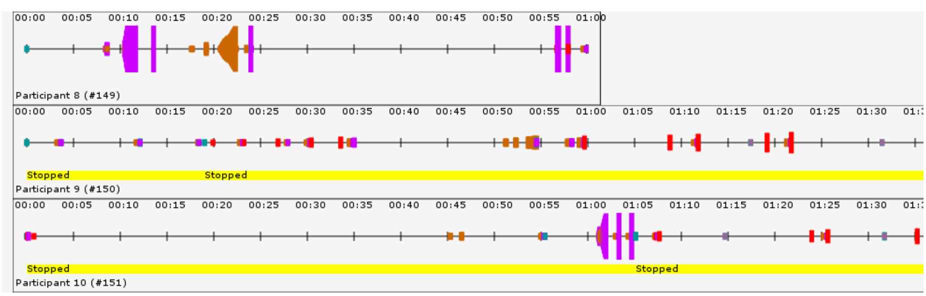


Fig. 3. Example of timelines generated by the tool in order to represent usage data

[3] http://www.w3schools.com/jsref/dom_obj_event.asp
[4] http://api.jquery.com/category/events/
[5] http://developer.apple.com/library/safari/#documentation/appleapplications/reference/safariwebcontent/HandlingEvents/HandlingEvents.html

Our proposal for representing event streams in timelines uses also the height of the time markers present in the timeline to represent the repetitions of a certain event in an element. This attribute is an indicator that allows evaluators to identify useless actions as well as to check repetitive mouse movements over bad designed link, misguided clicks on a non-clickable element, among others.

When visualizing the timelines of a certain task it is possible to zoom in and out in such representations and visualize them at the very basic event level or at the categories level. In the case of the categories visualization, the timeline uses different colors for different sets of events.

5 An Example Application

In this section we illustrate an application of our tool and show how an evaluator can infer usability issues from the visualizations provided by the tool. We report on a usability test of the Mobipocket[6] Web site, a known Web site for buying e-books. The tasks to be performed at the Mobipocket Web site, which were specified in the configuration of the WUP, were the following:

1. 'Search for books written by Jules Verne. Then, at the result list, add first Jules Verne's book to the cart.'
2. 'Browse the author list to check books available written by Umberto Eco.'
3. 'Access the cart and proceed to checkout. You don't have to login nor insert credit card information.' This task was dependent on the first task.
4. 'Find the contact form in order to send a message to the website. You don't have to send a message. Just reach the contact form.'

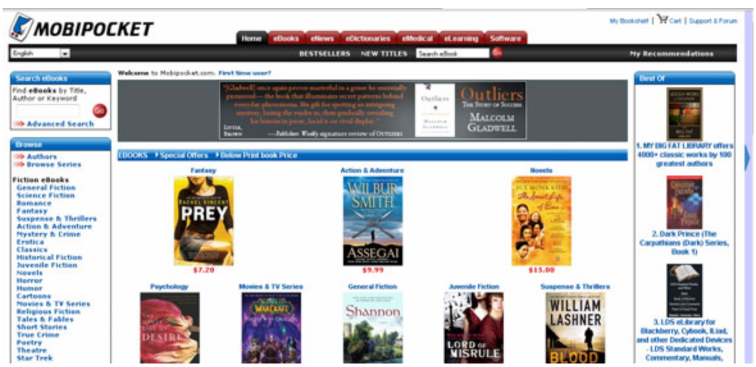


Fig. 4. The application considered in the case study

This usability evaluation involved 10 users. All of them were used to use Internet and were potential buyers of e-books, thus making part of the set of target users. Examples of results obtained from the timelines are the following:

[6] http://mobipocket.com

- During the task of adding a book of a certain author to the cart, the Mobipocket Web site showed as the first result a promotion of a packet of 100 authors. Figure 5 (timeline A) presents that one user was in doubt on clicking on the first result, moved the mouse over the result list and finally clicked on the correct book related to the task. It was possible to verify this issue by analyzing the timelines and checking that one participant performed the task in a different way from other participants. Thus, after zooming in the indicated timeline region it was possible to check details about events, for instance, mouse enter and mouse leave events over the first result of the search (indicated by the generated *id*) and then mouse enter event followed by a click on the add to cart button at the second result;

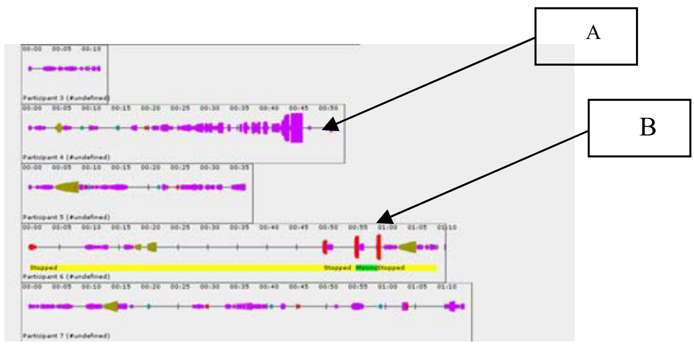


Fig. 5. Visualization of timelines; A) Repeated mouse move events indicating the doubt of the user in selecting the first search result; B) Resize events triggered by the popup blocker

- When adding a book to the cart, the feedback presented to the user was via a popup informing "This e-Book has been added to your shopping cart". Figure 5 shows page resizing events just after adding the book to the cart. After noticing this result in the timeline and mapping back to the UI it was possible to verify that when some browser's popup blocker message was presented, a sequence of resize events was then triggered;
- For the browsing authors list task it was possible to identify that some users tried to find 'Umberto Eco' under the letter 'U' page instead of the letter 'E' page, indicating lack of user guidance referring to the index, since users found unclear if the letter index refers to names or surnames;
- Considering the task of accessing the contact form we detected one usability issue: when users try to access the contact form via the 'contact us' link located at the footer of the Web site, they are sent to a Web page with another link to the contact page. This link in turn is an anchor to forum topics located at the middle of the contact Web page, making hard to users to find the contact form at the top of the page. Figure 6 presents the scrolling and mouse movements to find the link and then the same after accessing the page with forum topics and the contact form.

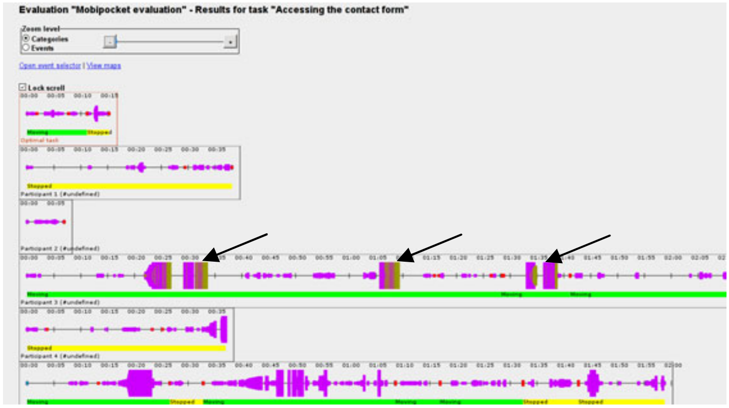


Fig. 6. Timelines referring to the access contact form. Arrows indicate mouse movements and scrolling deviation from other timelines.

6 Conclusions and Future Work

This paper presents WUP, a tool that allows evaluators to decide what tasks users should perform, and gather many types of data related to user interaction. The tool provides some graphical representations, which allow evaluators to analyze the data collected from a usability perspective. WUP also allows the end users to freely access the Web applications with any browser-enabled device without any constraint regarding where and when to perform such accesses during the test sessions. The case study reported indicated that the visual reports provided by the tool in form of timelines summarize event streams and highlight useless actions, allowing evaluators to identify usability problems, easing the task of mapping back events to actual actions occurred during user sessions. The experiment reported provided us with encouraging feedback, even if more validation will be carried out in the near future.

As future work we plan to extend the application of the Web Usability Probe considering the mobile Web, exploiting the flexibility in the definition of the events vocabulary in order to consider events related to mobile technologies.

Acknowledgments. We would like to thank FAPESP (grant #2009/10186-9) and all participants.

References

1. Google: Google Analytics (2009), http://www.google.com/analytics/index.html
2. Hartson, H.R., Castillo, J.C., Kelso, J.T., Neale, W.C.: Remote Evaluation: The Network as an Extension of the Usability Laboratory. In: CHI 1996, pp. 228–235 (1996)
3. Hilbert, D.M., Redmiles, D.F.: Extracting usability information from user interface events. ACM Comput. Surv. 32(4), 384–421 (2000)
4. Ivory, M.Y., Hearst, M.A.: The state of the art in automating usability evaluation of user interfaces. ACM Comput. Surv. 33(4), 470–516 (2001)

5. Malý, I., Slavík, P.: Towards Visual Analysis of Usability Test Logs Using Task Models. In: Coninx, K., Luyten, K., Schneider, K.A. (eds.) TAMODIA 2006. LNCS, vol. 4385, pp. 24–38. Springer, Heidelberg (2007)
6. Paganelli, L., Paternò, F.: Tools for Remote Usability Evaluation of Web Applications through Browser Logs and Task Models. Behavior Research Methods, Instruments, and Computers 35(3), 369–378 (2003)
7. Propp, S., Forbrig, P.: ViSE - A Virtual Smart Environment for Usability Evaluation. In: HCSE 2010, pp. 38–45 (2010)
8. Santana, V.F., Baranauskas, M.C.C.: A Prospect of Websites Evaluation Tools Based on Event Logs. In: Human-Computer Interaction Symposium. IFIP, vol. 272, pp. 99–104 (2008), doi:10.1007/978-0-387-09678-0_9

A Demo of a Dynamic Facial UI for Digital Artists

Pedro Bastos[1], Xenxo Alvarez[2], and Veronica Orvalho[2]

[1] Instituto de Telecomunicações, Faculdade de Engenharia da Universidade do Porto, Rua Dr. Roberto Frias, s/n 4200-465 Porto Portugal
[2] Instituto de Telecomunicações, Faculdade de Ciências da Universidade do Porto, Rua do Campo Alegre, s/n, 4169-007 Porto Portugal

Abstract. Character facial animation is difficult because the face of a character assumes many complex expressions. To achieve convincing visual results for animation, 3D digital artists need to prepare their characters with sophisticated control structures. One of the most important techniques to achieve good facial animation is to use facial control interfaces, also called facial user interfaces, or facial UI's. But facial UI's are usually dull and often confusing, with limited user interaction and no flexibility. We developed a concept and a working prototype of a dynamic facial UI inside the Blender [1] open-source software to allow their large community of digital artists to better control and organize the facial animation of a character. Our interactive system is running stable in the latest version of Blender and we started to build a full-face dynamic UI to show its interactive potential in a character's face.

Keywords: Accessibility and Usability, Novel User Interfaces and Interaction Techniques, Adaptive Interfaces, Dynamic Interfaces, Facial Interfaces, Facial Rigging and Animation, Digital Artists.

1 Introduction

Facial UI's [2] are GUI's (graphical user interfaces) with panels and controls that are used to facilitate the animation of a character's face. Facial UI's are usually placed side-by-side to the face of a 3D character in order for the animator to have easy access to its controls. The panels of a facial UI have a fixed position in space and generally each panel has a controller centered in the panel. The 2D spatial transformation of a control in a given panel affects the transformation of a corresponding locator in the skeleton of the character's mesh. The controls in the facial UI allow manipulating the character rig structure (the skeleton's bones) which in turn deforms the mesh.

The number of panels in the facial UI usually corresponds to the number of facial regions that need to be animated. The controls are grouped based on the anatomical areas of the face. There are references to help the digital artist build the animation controls for the face, such as the Facial Action Coding System [3].

The main benefit of a facial UI is the fact that it facilitates the understanding of the facial controls a character needs to have in order to simulate facial expressions [4] accurately. A facial UI enables the digital artist to interact with 2D controls that in fact report to 3D controls which are rigged to the mesh of the face of the character.

In essence, facial UI's prevent the visual confusion of the usual controls laid on top of the skin of the character. They also allow the user to have a better understanding of

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 358–359, 2011.

how each control is used, because the controls are constrained to the UI's panels. This prevents digital artists from doing unexpected manipulations of the controls which could result in awkward skin deformations and inaccurate facial animation.

But the common facial UI's also have considerable limitations. For instance, they occupy fixed positions on space. Also, the digital artist responsible for doing the UI generally places it side-by-side to the face of the character and the animator cannot change this. Also, facial UI's offer pre-established layouts with no flexibility for re-organization and re-dimensioning. This constrains and disorients the animator.

In general, a facial UI is a very complete system but is not straight forward enough nor it is the most user-friendly method, because the interactive abilities it offers are insufficient when compared to the flexibility of controls placed in the skin.

2 Our Solution: A Dynamic Facial UI

We developed a concept and a prototype of a dynamic facial UI in the Blender [1] software to help the large community of Blender digital artists do better animations for their characters faces. Our method improves the interaction of digital artists with facial UI's. It is an innovative technical solution for the 3D digital artist based on an extension of the features found in common facial UI's.

We already have a working prototype with very interesting features, such as the ability for the animator to easily relocate and re-dimension the panels in a facial UI. The animator can customize the limits and rearrange the layout of the facial UI, gaining more control over the animation of the several facial regions. It is possible to re-position the entire UI and/or a specific panel, transforming it to fit in the facial region as desired. The option to configure each panel provides more 2D space for manipulating the controls and a much better visualization and interaction.

Our prototype is able to join the abilities of the controls placed over the skin of the character to the abilities found in common facial UI's. With our prototype the digital artist can animate the face using the best of both worlds: having the flexibility of the controls placed directly in the skin and the benefits of using a facial UI.

Our concept has rendered us an article in the BlenderArt magazine. But we haven't yet had the opportunity to show it as a running demo. We would like to show it to an audience for the first time in the Interact2011 conference, with new improved functionalities and with the prototype applied to the several regions of the face.

We are also working to fully automate the process and make it compatible with any facial style the digital artist may need to animate. We attach a video to the submission of this document showing our results at the current development stage.

References

1. Blender Foundation, http://www.blender.org
2. Osipa, J.: Stop Staring: Facial Modeling and Animation Done Right. Sybex, Wiley Pub., Indianapolis, Indiana, USA (2010)
3. Ekman, P., Friesen, W.: The Facial Action Coding System. Research Nexus. Network Information Research Corporation, Salt Lake City, USA (2002)
4. Faigin, G.: The Artist's Complete Guide to Facial Expressions. Watson-Guptill Pub., New York (1990)

A Set of Customizable Games Supporting Therapy of Children with Cerebral Palsy

Beant Dhillon, Areti Goulati, Ioannis Politis,
Agata Raczewska, and Panos Markopoulos

Eindhoven University of Technology, Den Dolech 2, 5612 AZ
Eindhoven, the Netherlands
{B.K.Dhillon,A.Goulati,I.Politis,
A.M.Raczewska,P.Markopoulos}@tue.nl

Abstract. This research explores the potential of tangible tabletop technology for motor skills training of children with Cerebral Palsy. Therapists have emphasized the importance of customization of therapy programs to the needs of each patient. Five customizable games for the TagTiles tangible interactive gaming board are presented. These games have been developed based on feedback from 11 therapists from two clinics in the Netherlands, 9 children with Cerebral Palsy as well as 14 healthy children. The design process and the potential of this solution are briefly outlined.

Keywords: Cerebral Palsy, Motor Skills, Rehabilitation, TagTiles, Therapeutic Games, Design.

1 Introduction

Cerebral Palsy (CP) is a term describing a set of neurological disorders that cause physical disability, mainly affecting body movements and muscle coordination. In additional to traditional means of therapy, tabletop interactive surfaces offer a lot of potential. The TagTiles board [1] is a programmable grid of RFID sensors, able to provide visual and sound feedback, which is triggered by objects with attached RFID tags. The use of TagTiles in training for children with CP has been explored in the past [2] resulting in a series of games that support the training of arm-hand skills.

However, that work did not address several requirements that relate to the use of the games in actual therapy. Interviews with therapists emphasized the importance of supporting customization to each child's needs, and flexibility: therapists need to interrupt games and switch to other ones fluently during a therapy session.

2 CP Game Design

Input collected over 4 sessions with 11 therapists and 9 children with upper extremities CP from 2 different Dutch clinics resulted in the design of five new games and accompanying objects. These train different hand movements; grasping, wrist supination, and finger, shoulder and wrist extension. Objects of varying shapes, sizes,

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 360–361, 2011.

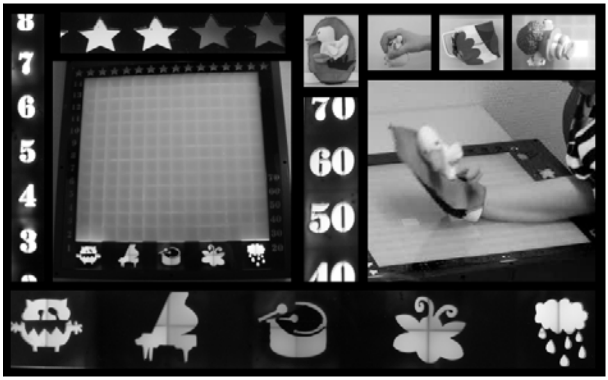


Fig. 1. The TagTiles board (left), the stars earned (top left), the game menu (bottom), the game speed indication (extreme left), the repetitions indication (middle), some of the objects designed (top right) and a child with CP interacting with the board (right)

textures and weights have been used along with adjustable Velcro tags (see Fig.1). All objects have been designed so as to prevent compensation.

The interaction in all games involves matching the flashing squares on the board with the objects in the child's hand. Every game uses an engaging background story (e.g., feeding a hungry monster or playing the piano). The therapists can control the difficulty level (i.e. game speed and number of repetitions) through a specially designed menu. To maintain motivation, encouraging feedback is provided through sounds and music. At the end of the game, feedback regarding performance is provided in terms of 'stars' earned by the child.

An evaluation session with 14 healthy children was performed as well in order to assess the very important fun aspect of the games. The games are now fully functional and ready to be embedded in a therapy setup.

References

1. Verhaegh, J., Fontijn, W., Hoonhout, J.: TagTiles: optimal challenge in educational electronics. In: 1st International Conference on Tangible and Embedded Interaction, pp. 187–190. ACM, New York (2007)
2. Li, Y., Fontijn, W., Markopoulos, P.: A tangible tabletop game supporting therapy of children with Cerebral Palsy. In: Fun and Games 2008: Second International Conference, pp. 182–193. Springer, Berlin (2008)

Mobile Total Conversation – Communication for All, Everywhere

Erik Zetterström

Omnitor AB,
Hammarby allé 93, 120 63 Stockholm, Sweden
erik.zetterstrom@omnitor.se

Abstract. This paper describes a demonstration of an open source Total Conversation client for the Android mobile phone platform. It also explains the rationale for Total Conversation and gives a brief overview of the open standards on which Total Conversation is based.

Keywords: Total Conversation, mobile, Deaf, Sign language, Video conference, Android, disabilities.

1 Introduction

Mobile phones have been a natural part of our life for, at least, the last 15 years. Mobile phones allow everyone the convenience to communicate with anyone anywhere, well not quite everyone after all. For those whose native language is a sign language this has not been the case. These individuals have been confined to using SMS or in some cases 3G video telephony. Not having equal access to telephony is clearly in violation of the United Nations Convention on the Rights of People with Disabilities [1], still the convention has been ratified by 99 countries [2]. This demonstration intends to showcase how mobile Total Conversation enables equal access to mobile telephony.

2 Total Conversation

The purpose of Total Conversation is to enable communication for all. Mobile phones and modern telecommunication networks have advanced to the point where it is now possible to deploy Total Conversation on most high end phones. The simultaneous use of audio, video and real-time text allows all users to find a suitable method for communication.

Video can be used for sign language or lip-reading for those who are deaf, hard of hearing, deaf-blind or late deafened.

Audio can be used for voice conversation according to the user's capabilities.

Real-time text can be used by everyone to convey information that is tedious to convey in other media. It can also be used to support other media in the call for example; through a captioned telephony service or in some cases to type a word instead of speaking it for someone with a speech impairment. For deaf blind users it could be used to receive text while the response may be sent in sign language.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 362–363, 2011.

Total Conversation is based on open standards and is defined in [3]. As in most open IP telephony systems today the Session Initiation Protocol (SIP) [4] and RTP [5] is used. H.263 [6] and H.264 [7] should be supported for video. H.263 is widely implemented and the newer H.264 offers better video quality with less bandwidth. G.711 [8] is recommended for audio since it is widely supported and offers good quality. Real-time text should be implemented by T.140 [9]. It provides character by character transmission, giving the user a sense of a real conversation.

For good communication in sign language the video resolution should not be below QCIF (176 × 144) and 20 frames per second (fps) is recommended [10].

3 Demonstration Description

The demonstration will showcase the benefits of Total Conversation being added to the open source mobile Total Conversation client linphone [11] running on Android mobile phones. Calls to and from this client to other proprietary and open source clients Total Conversation will demonstrate the importance of open standards for interoperability.

During the demonstration lessons learned will be presented, for example regarding CPU and network requirements.

References

1. Convention on the Rights of Persons with Disabilities, `http://www.un.org/disabilities/default.asp?id=150`
2. Convention on the Rights of Persons with Disabilities - Convention and Optional Protocol Signatures and Ratifications, `http://www.un.org/disabilities/countries.asp?id=166`
3. ITU-T F.703 Telecommunications accessibility guidelines for older persons and persons with disabilities, `http://www.itu.int/rec/T-REC-F.703`
4. IETF RFC3261 SIP: Session Initiation Protocol, `http://www.ietf.org/rfc/rfc3261.txt`
5. IETF RFC3550 RTP: A Transport Protocol for Real-Time Applications, `http://www.ietf.org/rfc/rfc3550.txt`
6. ITU-T H.263 Video coding for low bit rate communication, `http://www.itu.int/rec/T-REC-H.263/`
7. ITU-T H.264 Advanced video coding for generic audiovisual services, `http://www.itu.int/rec/T-REC-H.264/`
8. ITU-T G.711 Pulse code modulation (PCM) of voice frequencies, `http://www.itu.int/rec/T-REC-G.711/`
9. ITU-T T.140 Protocol for multimedia application text conversation, `http://www.itu.int/rec/T-REC-T.140/`
10. ETSI EG 202 670 Human Factors (HF) User Experience Guidelines for real-time communication services expressed in Quality of Service terms
11. linphone – an open source video SIP phone for desktop & mobile, `http://www.linphone.org/`

Storytelling Meets the Social Web: An HTML5 Cross-Platform Application for Older Adults

Tiago Boldt Sousa[1], Pedro Tenreiro[1],
Paula Alexandra Silva[1], and Eduarda Mendes Rodrigues[2]

[1] Fraunhofer Portugal, Rua Alfredo Allen 455/461 4200-135 Porto Portugal
[2] Departamento de EngenhariaInformática, Faculdade de Engenharia da Universidade do Porto, Rua Dr. Roberto Frias, s/n 4200-465 Porto Portugal

Abstract. This demonstration presents a storytelling application specifically designed for older adults to share stories and thoughts. Studies claim that older adults commonly have difficulties in engaging with on-line social networks [1], but increased social inclusion and sense of well-being has been observed in those who engage [2]. While following a user-centered design approach, we have developed an HTML5 device-independent and intuitive social web application which addresses older adults' specific needs and age-related impairments, allowing them to connect to their friends and family through storytelling.

Keywords: Storytelling, Online Social Networks, User-Centered Design, Older Adults, HTML5.

1 Introduction

Unlike a few decades ago, today's households no longer have more than two generations [4], resulting in older adults living by themselves. Distance from family can result in feelings of loneliness and can limit their ability for sharing their legacy, something which is paramount for them. Furthermore, transportation and communication problems are known constraints due to the decline of their physical and mental conditions, resulting in a general decline of their well-being and happiness [1]. While they can be positively influenced by the use of online social networks [2], younger people are the mainstream audience of these services [3]. Feature-rich services and web pages with large amounts of information are common and difficult to use by those lacking computer skills; this results in a general inhibition towards the exploration of technology from the elderly. To address these problems and to fight the social exclusion of older adults, we developed a prototype of a storytelling applicationthat leverages an existing onlinesocial network (Facebook) for connecting older adults with their relatives.

2 Cross-Platform Storytelling on the Social Web

2.1 Application Design

The prototype application was implemented using a set of novel web technologies commonly denominated as HTML5, taking into account special design constraints

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 364–365, 2011.

identified through user-centered design, which highlighted common impairments and difficulties of older adults' interaction. Several observations and usability tests were pursued with a sample of 15 older adults to inform the design. We have used the test results to accommodate a small learning curve, as well as a pleasant experience for any older adult user. Those studies provided us with our main goals which were to enable text and image sharing, as well as to allow follow-up comments on them. The inclusion of meta-data in each story (for example the user's mood or location) appeared as interesting, but secondary.

2.2 Prototype Implementation

The implementation of this storytelling application followed a user-centered design approach, in order to create an easy to adopt application by senior users. This demonstration aims at showing how the application enables older adults to share stories through online social networks, specifically Facebook. Our target users consist of older adults without any prior experience with information and communication technologies. As such, the client application only provides access to a small subset of all Facebook's features, those identified as the most relevant for the older adults, while also keeping the application easy to use. Despite primarily aiming at older adults, any individual with a Facebook account will be able to use the storytelling application as well as to interact with story-tellers through the Facebook service itself. The prototype uses elements from the recent HTML5 specification, such as localStorage for keeping local data. This guarantees seamless portability between compatible browsers, enabling the user to choose her preferred device to use the application.

3 Conclusions and Future Work

Our application provides an alternative communication channel between olderadults and their friends and family, hoping to increase their well-being throughthe elimination of feelings of loneliness.

Due to the fast-paced evolution of the HTML5 standard, extra features mightbe included in the project before the date of the demonstration. While sharingtext and images are the initial features, video and audio capture should also beintegrated into the application as soon as they are available in web browsers(already described in the current HTML5 specification).

References

1. Chen, Y.: Usability Analysis on Online Social Networks for the Elderly (2009)
2. Gibson, L., Moncur, W., Forbes, P., Arnott, J., Martin, C., Bhachu, A.S.: Designing Social Networking Sites for Older Adults (2010)
3. Pingdom: Study: Ages of Social Network Users (2010)
4. Williams, B.K., Sawyer, S.C., Wahlstrom, C.M.: Marriages, Families & Intimate Relationships. Pearson, Boston (2005)

Tablexcel: A Multi-user, Multi-touch Interactive Tabletop Interface for Microsoft Excel Spreadsheets

Guillaume Besacier[1,2]

[1] LIMSI-CNRS, BP 133, 91403 Orsay Cedex, France
[2] Université Paris-Sud, 91405 Orsay Cedex, France
guillaume.besacier@limsi.fr

Abstract. In this paper, we present Tablexcel, a tabletop interface to Microsoft Excel. Single-user, desktop-based computer applications are pervasive in our daily lives and work. An application like Microsoft Excel, a widely deployed spreadsheet application, is used by a large number of businesses and users. Often, several users will collaborate on the creation of a spreadsheet, for example exchanging Excel files by e-mail. A multi-user, multi-touch interactive tabletop could create better working conditions, but Excel is not compatible with tabletop interfaces. Tablexcel use the scripting capabilities of Excel to extract live data from Excel files, and display them in a tabletop-appropriate way. Multiple users can interact with the Tablexcel interface using tabletop interactions, like gestures or rotating windows. Tablexcel manage the collaborative aspect of the interaction and send the resulting modifications to the original Excel application, which update the formulas, graphs, macros, etc.

Keywords: Interactive tabletop, legacy application, spreadsheet, scripting.

1 Introduction

Single-user, desktop-based computer applications are pervasive in our daily lives and work. Being able to use these applications with novel interactive systems, like multi-touch tabletops, is a requirement for a large deployment of these systems in a production environment. Indeed, these applications, whether they are widely used or specialized business applications, are essential to their users.

Microsoft Excel, a widely deployed spreadsheet application, is used by a large number of businesses, either for its own value or as a platform for third-party applications programmed with its macro language. Often, several users collaborate to create a single spreadsheet. This author had received and sent back Excel file by e-mail for such diverse purpose as cooperatively making a planning, drafting a list, or filing forms. Excel files are also edited during face-to-face meeting, for example to tweak the weight given to the various grades during an academic jury. Often, a meeting leader will modify the master file on his or her laptop, while the other participants try to follow and make the same modification on their own laptop. A multi-user, multi-touch tabletop could create better working conditions in both situations.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 366–369, 2011.

Researchers have created new applications, designed for interactive tabletops, to fill the role of existing applications. For example, a spreadsheet application [4] was programmed with the T3 tabletop toolkit. But it is not compatible with existing file formats, and offer only basic functionalities. In addition, it doesn't offer a rich, collaborative multi-user experience, as it uses an off-the-shelf single-user component.

This lack of real-world applications limits the appeal of tabletop computing for businesses and end-users, but it also limits our ability, as HCI researchers, to conduct long term studies and field studies on tabletop usage. The only such study we are aware of was conducted using a single user in a desktop environment using a tabletop to emulate a mouse [5].

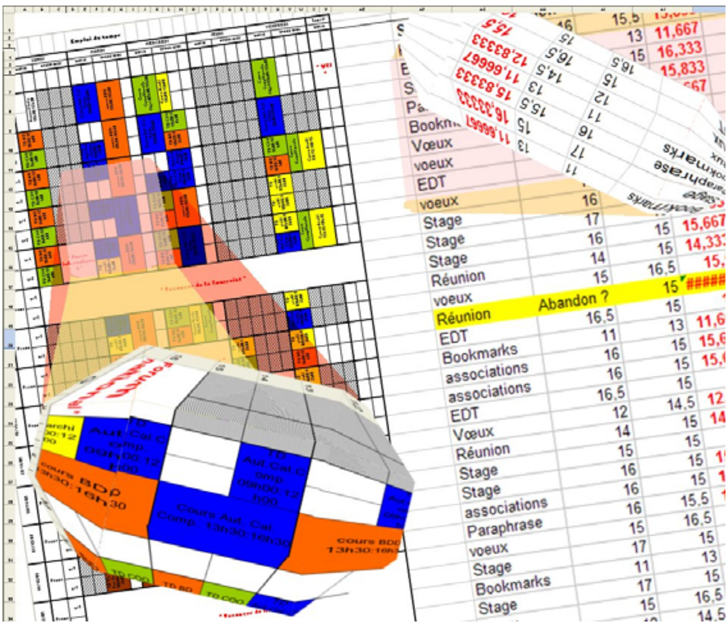


Fig. 1. Prototype of the Tablexcel user interface. Each user can have their own focus point.

2 Implementation

Several technologies have been indentified to use existing applications with innovative interactive systems [1]. In this paper, we describe our implementation of Tablexcel, a tabletop interface to Microsoft Excel. Excel exposes a scripting interface. Its purpose is to facilitate tasks automation and applications interactions. For example, a third-party a word processing application might ask a spreadsheet application for some data to create a mailing. Scripting allows Tablexcel to access Excel raw data without formatting for a particular viewport, with the burden of presenting the data in a visual way delegated to our interface.

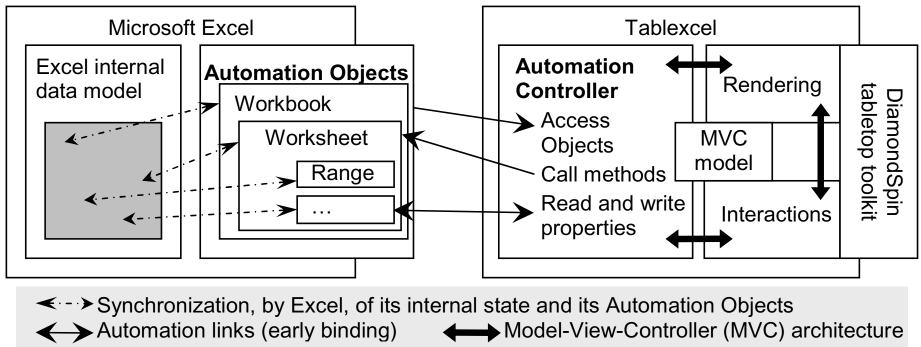


Fig. 2. Tablexcel MVC architecture delegates the Model part to the actual Excel application

Excel exposes its data model with Automation objects. Automation is an inter-process communication mechanism based on the Microsoft Component Object Model (COM). It provides an infrastructure whereby applications called Automation Controllers can access and manipulate (i.e. set properties of or call methods on) shared Automation Objects that are exported by other applications [2]. The Excel Automation Object Workbook represents a file. It contains Worksheet objects, which contain Range objects representing a cell. Each cell has several properties: the numeric value of the cell, the formula used to compute it, its size, font, color, etc.

We implemented an Automation Controller, called Tablexcel, in C++ and Java, which uses the Automation Objects exported by Excel as its data model (Figure 2). Tablexcel uses the Model-View-Controller (MVC) architecture. The Model is composed of Automation Objects exported by Excel. Excel synchronizes theses objects with its internal state. The Controller and View are implemented using the DiamondSpin tabletop toolkit [3]. For maximum speed, the rendering is done in OpenGL using a custom algorithm created specifically to display Excel spreadsheets.

3 Interacting with Tablexcel

Tabletop spreadsheets prototypes developed so far have retained the desktop metaphor, using windows, scrollbar and similar widgets to interact with their users. We think this metaphor is not the most efficient when several users are working on the same spreadsheet. Our inspirations for Tablexcel were the interactions one can have with a large map (which is an area frequently studied on tabletops).

The spreadsheet is displayed on the whole table. Is can be translated and rotated with a finger touch, and zoomed with a pinch gesture. The translation and rotation use a kinetic relation between the finger movement and the spreadsheet movement rather than a direct mapping: the movement starts slowly and then pick up speed. As the spreadsheet is shared by all the users, it can disturb a user work if another user suddenly moves it. A slow initial speed allows for easy cancellation of the movement in its initial phase. A user can cancel the movement by putting their whole hand on the table to metaphorically hold the spreadsheet in place.

This view of the spreadsheet cannot be modified. The size and resolution of most current tabletops doesn't allow small enough widgets to interact directly with this view, which need to be broad enough that all the users' centers of interest fit in the tabletop surface. Instead, users can open an interactive secondary view (two such views are opened in Figure 1), by delimiting an area of interest between the side of their hands. These views are then managed like windows in other tabletop applications: they can be rotated, moved, zoomed, etc. Each user has their own focus, and a formula bar on the tabletop border next to them (not visible in Figure 1). The focused cell is editable using a physical wireless keyboard per user.

4 Conclusion

We presented Tablexcel, a multi-user, multi-touch tabletop interface to Microsoft Excel using the scripting capabilities of Excel to extract live data from Excel files. In future works, we plan to use Tablexcel to study users' interactions with an interactive tabletop in field studies using users' actual data and work processes.

References

1. Besacier, G., Vernier, F.: Toward user interface virtualization: legacy applications and innovative interaction systems. In: Proc. EICS 2009, pp. 157–166 (2009)
2. Kruglinski, D., Wingo, S., Shepherd, G.: Automation. In: Programming Microsoft Visual C++ 6.0, 5th edn. Microsoft Press, Redmond (1998)
3. Shen, C., Vernier, F., Forlines, C., Ringel, M.: DiamondSpin: An Extensible Toolkit for Around-the-Table Interaction. In: Proc. CHI 2004, pp. 167–174 (2004)
4. Tuddenham, P., Robinson, P.: T3: Rapid Prototyping of High-Resolution and Mixed-Presence Tabletop Applications. In: Proc. Tabletop 2007, pp. 11–18 (2007)
5. Wigdor, D., Perm, G., Ryall, K., Esenther, A., Shen, C.: Living with a Tabletop: Analysis and Observations of Long Term Office Use of a Multi-Touch Table. In: Proc. Tabletop 2007, pp. 60–67 (2007)

Assessing Short-Term Human-Robot Interaction in Public Space[*], [**]

Jakub Złotowski

HCI and Usability Unit, ICT&S Center, University of Salzburg,
Sigmund-Haffner-Gasse 18, 5020 Salzburg, Austria
{jakub.zlotowski,manfred.tscheligi,astrid.weiss}@sbg.ac.at

Abstract. This thesis discusses an evaluation approach for assessing social acceptance of short-term HRI in public space with special emphasis on robots seeking information from pedestrians.

Keywords: Human-Robot Interaction, Evaluation Methodology, Measures and Metrics, Social Acceptance.

1 Context and Motivation

In recent years the research on Human-Robot Interaction (HRI) started moving from highly-controlled laboratory experiments to field studies in natural human environments. However, despite increased interest in spontaneous and time limited (short-term) HRI, due to lack of common benchmarks and theoretical concepts being evaluated by researchers, it is not possible to compare the results of these studies [4]. Moreover, the context of public space is challenging for conducting reliable studies, as new methods are needed in order to address the problem of less controlled environment compared to laboratory studies [2, 9].

There is little doubt that successful integration of robots in human environments can accelerate the speed of further development of that technology. High usability of robots is required so that people will be eager to use them. Furthermore, there are also two other key aspects which must be addressed in order to accelerate that process: social acceptance and user experience. HRI, as a relatively young discipline, suffers from lack of reproducibility of studies and comparability of results since currently they are often platform specific. However, Bartneck [1] pointed out that the field reached the stage when it is necessary to conduct studies in a consistent manner in order for it to progress further.

The work presented in this paper was developed and is being evaluated as part of the EU funded FP7 project Interactive Urban Robot (IURO). The scenario explored in this project involves development of a robot that is capable of navigating in densely populated human environments using only information obtained from encountered humans. As an alternative to GPS enabled and Internet access provided robots, the IURO must find its way to a destination located in another part of a city by asking

[*] Supervisor: Manfred Tscheligi
[**] Advisor: Astrid Weiss

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 370–373, 2011.

pedestrians for directions. A robotic system must be socially accepted so that it receives help from strangers. Therefore, it is necessary to create a platform independent set of benchmarks and methods suitable for evaluation of robots which are interacting with bystanders.

2 Background and Related Work

Human-Computer Interaction (HCI) and HRI have provided several evaluation methods to assess various aspects of human interaction with social robots. However, there has been relatively little work done which aims at providing a holistic evaluation approach for short-term HRI. Weiss [12] proposed an evaluation framework for human-robot collaboration that included four main factors: Usability, Social acceptance, User experience and Societal impact (USUS). Each of these factors was then operationalized by several indicators, which were linked with suitable research methods for addressing them.

In many aspects HRI in public space differs from interaction in other areas investigated in the field. Usually such interaction is only short-term and people spend limited time with a robot [10]. Therefore, their ability to adapt to a robot is reduced. Furthermore, since interaction in public space is spontaneous, there is no extensive preparation on a person's side, such as reading manual. As a result using natural human interaction channels is critical for robots operating in public space [10]. In addition, specific aspects of this context require adapting HRI evaluation frameworks and providing new methods for addressing them during studies in natural human environments. As Weiss [11] noted the USUS evaluation framework needs to be adjusted for different contexts of HRI.

Human beings are a reach source of information in rapidly changing environments, such as public space. Since modern robots do not have human capabilities yet, and even if they had, it will be often necessary for a robot to approach strangers in order to obtain information required for successful accomplishment of various tasks. This will be only possible if people accept robots as social actors.

There is strong evidence under the Computers are Social Actors paradigm that human interaction with technology is basically social, e.g. [3, 7]. Furthermore, compared with other interactive technologies humanoid robots encourage more anthropomorphic mental models [5]. Therefore, it seems appropriate to expect that people providing help to robots will base their decision on similar criteria as when other human requests help. Latané and Darley [6] proposed the Five Stage Model of Helping, which a bystander goes through before help is provided for a person in need. Each stage represents an important decision and if at any point the outcome is negative, it will result in no help being provided. The stages proposed by them are:

1. Noticing the situation
2. Labeling the situation correctly as emergency
3. Assuming responsibility for helping
4. Deciding how to help
5. Implementing the decision

Furthermore, they identified factors, which can influence the decision at various stages. The model begins before the interaction is initiated (pre-beginning sequence [8]) and ends when help is being provided. Therefore, if pedestrians' mental model will be similar for human strangers and robots alike, it would be possible for robots to impact people's decision for providing help.

3 Research Hypothesis and Methods

My assumption is that if the Five Stage Model of Helping is also applicable for HRI it should be possible to increase a robot's chances of receiving help from strangers by addressing critical benchmarks on each of the stages. Therefore, the main goal of my dissertation is to create an evaluation approach for assessing social acceptance of short-term HRI in public space with special emphasis on robots seeking information from pedestrians. It should constitute of two parts: theoretical and methodological.

The theoretical part of the evaluation approach is based on the described above Five Stage Model of Helping [6]. However, compared with a human being, usability of a robotic system can affect whether help will be provided completely and successfully. Hence, an additional sixth stage (Process of helping) has been added to this model. This extended framework called the MOdel of HElping Robots (MOHER) should allow evaluation on all stages from noticing the situation until a person finishes providing help to a robot. Robot benchmarks, which are important in order to increase probability of bystander passing to the next stage, are assigned to each stage of the MOHER. There are multiple indicators identified at each stage and the same benchmark can be important at various stages.

The USUS evaluation framework [11] is used as the starting point for identifying social acceptance indicators. Only benchmarks which are meaningful in the new context are used and some of them had to be redefined. Furthermore, based on broad literature review from Psychology, HCI and HRI new indicators were added. However, it is not possible to obtain full picture of short-term HRI without analyzing the relevance of two other factors: usability and user experience. Thus, they are also incorporated in the MOHER.

The methodological part provides means for the evaluation of the MOHER. An evaluation plan has been created and each stage of the model is explored in a dedicated case study. The methods used include qualitative and quantitative studies, such as: field trial, heuristic evaluation or focus troupe. In addition to existing HRI practices, various methods from HCI and Ubiquitous computing, such as Co-discovery or Bluetooth device scanning, will be utilized in order to address the theoretical concepts. The first video-based case study on approach trajectories with 30 participants has just been conducted and the results are promising for feasibility of the proposed evaluation approach. Moreover, SoAc and UX questionnaires developed by Weiss [11] will be adjusted and validated in the IURO project's scenario in order to learn more about HRI in public space.

4 Expected Contributions

Based on these 6 case studies, guidelines, which will indicate what methods are suitable for different stages and what benchmarks should be evaluated, will be created for fellow researchers. As a result, it is believed that the evaluation process of short-term HRI in public space will be simplified and identification of the stage at which a robot's performance has to be improved will be aided. The PhD research should also improve reproducibility of studies as the same benchmarks can be used for various studies and the results could be compared.

As an exploratory work, the conducted case studies will not be able to guarantee holistic evaluation of the concept. However, identified stages, benchmarks and validated methods for evaluating them, together with the guidelines for other evaluators can have positive impact on HRI design process.

References

1. Bartneck, C., Kulić, D., Croft, E., Zoghbi, S.: Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. International Journal of Social Robotics 1, 71–81 (2009)
2. Dix, A., Finlay, J., Abowd, G.: Human-computer interaction. Prentice Hall, Englewood Cliffs (2004)
3. Fogg, B., Nass, C.: How users reciprocate to computers: an experiment that demonstrates behavior change. In: CHI 1997 Extended Abstracts on Human Factors in Computing Systems: Looking to the Future, CHI EA 1997, pp. 331–332. ACM, New York (1997)
4. Fong, T., Nourbakhsh, I., Dautenhahn, K.: A survey of socially interactive robots. Robotics and Autonomous Systems 42(3-4), 143–166 (2003)
5. Kiesler, S., Hinds, P.: Introduction to this special issue on human-robot interaction. Hum.-Comput. Interact. 19, 1–8 (2004)
6. Latané, B., Darley, J.: The unresponsive bystander: why doesn't he help? Prentice Hall, Englewood Cliffs (1970)
7. Lee, E.J., Nass, C., Brave, S.: Can computer-generated speech have gender?: an experimental test of gender stereotype. In: CHI 2000 Extended Abstracts on Human factors in Computing Systems, CHI EA 2000, pp. 289–290. ACM, New York (2000)
8. Schegloff, E.: On the organization of sequences as a source of "coherence" in talk-in-interaction. Conversational Organization and its Development 38, 51–77 (1990)
9. Tanaka, F., Movellan, J.R., Fortenberry, B., Aisaka, K.: Daily hri evaluation at a classroom environment: reports from dance interaction experiments. In: Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction, HRI 2006, pp. 3–9. ACM, New York (2006)
10. Thrun, S., Schulte, J., Rosenberg, C.: Interaction with mobile robots in public places. IEEE Intelligent Systems, 7–11 (2000)
11. Weiss, A.: Validation of an evaluation framework for human-robot interaction. The impact of usability, social acceptance, user experience, and societal impact on collaboration with humanoid robots. Ph.D. thesis, University of Salzburg (2010)
12. Weiss, A., Bernhaupt, R., Lankes, M., Tscheligi, M.: The USUS evaluation framework for human-robot interaction. In: AISB 2009: Proceedings of the Symposium on New Frontiers in Human-Robot Interaction. Citeseer (2009)

Barefooted Usability Evaluation: Addressing the Mindset, Resources and Competences*

Anders Bruun

Aalborg University,
Dept. of Computer Science,
Selma Lagerlöfs Vej 300,
9220, Aalborg Ost, Denmark
bruun@cs.aau.dk

Abstract. Lack of usability specialists, high resource requirements and developer mindsets are three considerable barriers for introducing usability engineering into software companies. This Ph.D. project explores the effect of letting software developers and end users apply usability engineering methods instead of a specialist, a solution which may reduce the barriers.

Keywords: Usability evaluation, Training, Software Developers, End Users.

1 Introduction

During the last decade software companies have increased their focus on introducing usability engineering into their development processes. However, one challenge experienced by companies is the sheer lack of usability specialists in the industry, which leads to missing competences and ultimately problems of incorporating usability [10]. However, small companies do not even have the privilege of staffing usability specialists as these have to cope with the constraint of low budgets. In practice this means that small software companies do not have the funds to pay for comprehensive consultancy or staffing of usability specialists [7] as they are expensive to hire [9]. Another known problem, denoted the "developer mindset" [1], regards attitudinal aspects such as missing acceptance or low prioritization from the side of the developers. This is also a hindrance to future usability work within a company. Thus, small software companies wanting to focus on usability engineering face these three barriers: Missing competences, high resource demands and developer mindset.

If usability specialists are in scarce supply and small companies cannot afford to hire these, who, then, could undertake the usability work? This project explores the solution of letting the developers themselves and the end users apply usability engineering methods instead of a specialist, which in turn may reduce the three barriers.

---

* This dissertation work is supervised by Prof. Jan Stage, Department of Computer Science, Aalborg University, Denmark, jans@cs.aau.dk

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 374–377, 2011.

2 Theoretical Foundation and Hypotheses

Before moving on, the following question should be answered: Why is it that, by letting developers and end users apply usability engineering methods, the barriers of missing competences, high resource demands and the developer mindset will be reduced? To answer this we may turn to Mao Zedongs cultural revolution in China in the 1960's. One vision behind this revolution was to bring better healthcare services to farmers living in rural areas. The problem at the time was the existence of an urban bias of doctors leading to scarce medical competences outside the cities. To counter this problem, Mao transferred medical knowledge from urban doctors to rural areas by letting farmers receive training in basic medicine such as the delivery of babies, how to ensure better sanitation etc. [6]. Thus, some of the farmers would work part time in the rice fields walking around barefooted and part time as doctors in the local area. According to the World Health Organization these so called "barefoot doctors" reduced the amount of resources required, which was observed as reductions in healthcare budgets. These savings were caused by the preventive and tertiary care provided [6]. An additional advantage was that rural farmers perceived the barefoot doctors as peers hereby respecting their advice [11].

Although finding ourselves in a domain differing considerably from Chinese healthcare, we today face a similar challenge as before Mao's Cultural Revolution. Caused by missing competences, high resource requirements and developer mindset, small software companies, in a usability sense, have become the modern rural areas not being able to focus on usability work. This parallel challenge may be met by a parallel solution, i.e. by providing basic training in usability engineering methods, developers and end users could potentially act as barefooted usability personnel. This obviously could reduce the barrier of missing competences and could potentially ease the problem of scarce supply of specialists. Additionally, resource requirements may be brought down as usability engineering is preventive in avoiding expensive projects failures [2] and because the need to employ usability specialists is lessened. Also, other software developers may perceive barefooted usability personnel as peers, hereby accepting the advice given and reducing the barrier of developer mindset.

Thus, based on the notion of barefoot doctors the following hypotheses are derived:

- **H1:** Barefooted usability engineering reduces the barrier of missing competences.
- **H2:** Barefooted usability engineering reduces the barrier of resource requirements.
- **H3:** Barefooted usability engineering reduces the barrier of developer mindset.

3 Research Questions

Up until this point the term "usability engineering" has been applied, which covers several methods for analysis, design and evaluation activities. During my Ph.D. I do not have time to train developers and end users in all methods, which is why I focus on training these in evaluation methods, and more specifically in conducting user based usability evaluations (as opposed to inspection based). User based evaluations have proven effective in convincing developers of the importance of usability work as they provide first hand insights in difficulties experienced by end users [8] and may, thus, ease the introduction of usability engineering into software companies.

The following two research questions are addressed in the thesis:

- **RQ1:** What are the previous experiences and future research needs on training developers and end users in conducting usability evaluations?
- **RQ2:** How do usability evaluations conducted by developers and end users help in reducing the barriers of missing competences, high resource demands and developer mindset? How do developers and end users perform in comparison?

No previous studies have examined how usability evaluations conducted by developers and end users help in reducing all of the three barriers. A recent literature review identifies future research needs regarding training of software developers. The review shows that empirical studies of training developers in user based evaluation methods in general are needed [4].

Additionally, previous studies have shown that evaluations conducted by end users are possible, e.g. by applying remote asynchronous methods without the presence of an evaluator, see e.g. [5]. However, these studies focus on users' performance versus specialists', which leaves room for comparing user performance with that of developers.

4 Research Method

In this project my aim is to intervene by training developers and end users in applying usability evaluation methods and to study how these subjects apply methods in a natural setting. These characteristics call for the action case as a research method [3]. To ensure data validity, multiple data collection methods will be employed:

- **Observation:** E.g. observations on correctness in conducting evaluations, amount of resources required (hours, money etc.) and how usability problems are prioritized and observations on product improvements.
- **Surveys and interviews:** For instance asking developers and other stakeholders how results from usability evaluations are used, how evaluations should be prioritized, what they perceive as barriers for doing more usability engineering etc.
- **Document analysis:** Analysis of usability reports with respect to no. of problems identified, quality of problem descriptions, reflections on the usability evaluation procedure etc.

5 Research Contribution

This project will contribute to the HCI research community by providing understanding of the barefooted theory as a driver for reducing the most significant barriers in incorporating usability engineering. This understanding emerges through reference studies in which the theory is applied to create changes in industrial practice. Thus, understanding and change are the types of outcomes this Ph.D. will bring to the research community and industrial practice respectively [3].

References

1. Bak, J.O., Nguyen, K., Risgaard, P., Stage, J.: Obstacles to usability evaluation in practice: a survey of software development organizations. In: Proc. NordiCHI 2008, pp. 23–32. ACM Press, New York (2008)
2. Bias, R.G., Mayhew, D.J.: Cost-Justifying Usability: An update for the Internet Age. Morgan Kaufmann, San Francisco (2005)
3. Braa, K., Vidgen, R.: Interpretation, intervention, and reduction in the organizational laboratory: a framework for in-context information system research. Accounting, Management and Information Technologies 9(1), 25–47 (1999)
4. Bruun, A.: Training software developers in usability engineering: a literature review. In: Proc. NordiCHI 2010, pp. 82–91. ACM, New York (2010)
5. Castillo, J.C., Hartson, H.R., Hix, D.: Remote usability evaluation: Can users report their own critical incidents? In: Proc. CHI 1998, pp. 253–254. ACM Press, New York (1998)
6. Daqing, Z., Unschuld, P.U.: China's barefoot doctor: Past, present, and future. The Lancet 372(9653), 1865–1867 (2008)
7. Häkli, A.: Introducing user-centered design in a small-size software development organization. Helsinki University of Technology, Helsinki (2005)
8. Høegh, R.T., Nielsen, C.M., Overgaard, M., Pedersen, M.B., Stage, J.: The Impact of Usability Reports and User Test Observations on Developers' Understanding of Usability Data: An Exploratory Study. Int. Journal of Human-Computer Interaction 21(2), 173–196 (2006)
9. Nielsen, J.: Usability inspection methods. In: Proc. CHI 1994, pp. 413–414. ACM Press, New York (1994)
10. Nielsen, J.: Finding usability problems through heuristic evaluation. In: Proc. CHI 1992, pp. 373–380. ACM Press, New York (1992)
11. Markle, W.H., Fisher, M.A., Smego, R.A.: Understanding Global Health. McGraw-Hill, New York (2007)
12. Scholtz, J., Laskowski, S., Downey, L.: Developing usability tools and techniques for designing and testing web sites. In: Proc. Conference on Human-Factors & the Web (1998)

Collaborative Human-Machine Communication: User-Centered Design of In-Vehicle Speech Dialog Systems

Linn Hackenberg

VOLKSWAGEN GROUP,
38436 Wolfsburg,
Germany
Department of Psychology, University of Brunswick, Brunswick, D
linn.hackenberg@volkswagen.de
Dissertation advisor: Prof. Dr. Mark Vollrath, University of Brunswick
mark.vollrath@tu-braunschweig.de

Research-area: User-centered Design, speech dialog systems, in-vehicle HMI

Description of the research topic: Evaluation of Speech Dialog Systems that make use of collaborative strategies from human conversations by providing continuous and appropriate feedback whilst showing dynamic interaction-structures.

1 Research Problem

One of the main challenges of today's speech dialog systems (SDS) is that they fail to provide appropriate feedback about status and capabilities of the system. Until now they lack any representation of comprehension. Immediate acknowledgements or back-channels, known from everyday conversation with people [1] are missing. This leads to high uncertainty of users about what the system is up to. Operation errors like talking too early or entering input again while the system is still processing are due to state confusion which demonstrates the users' need for process indicators. These operating errors consequently lead to recognition errors followed by a low user satisfaction. In addition, there is substantial evidence that counter-intuitive system interaction leads to high cognitive workload [2], which is a severe problem especially in the in-vehicle dual-task context.

Another challenging issue is the perceived length of dialogs. Inflexible interaction structures of current systems lead to low user satisfaction with the efficiency of in-vehicle SDS. Analogous to Grice's maxim of manner [3] the least long dialog is preferred. Rather than the number of words used per speech output, length here refers to situation-appropriate system responses. In particular, dispensable dialog loops demanding the user to confirm his input are often misinterpreted as recognition errors and eventually lead to poor ratings of the system. Moreover, these redundant confirmations provoke specific operating errors (talking to early), hyperarticulation and frustration on the part of the users. Latter goes back to a mismatch with existing mental models of interpersonal communication. Here the amount of effort that both

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 378–381, 2011.

partners dedicate to a dialog step before moving on is determined by their grounding criteria [4] and is highly context-sensitive. The higher the grounding criterion is set, the more evidence conversational partners will require before concluding that an utterance is accepted [5]. The problem postulated here is that the systems grounding criterion is inflexible and so at most of the turn-takes too high.

Until now, the process of grounding, which entails people systematically seeking and providing evidence about what has been said and understood does not sufficiently take place with SDS. It can be shown that this lack of grounding reduces the intention to use and the joy-of-use experience. Therefore, transparency is an extremely important precondition in order to create user acceptance for new speech interfaces.

Traditional approaches mainly focus on improving the correctness of the underlying speech recognition. The presented experiments employed here establish dialog strategies from human conversation to investigate whether these strategies can reduce the amount of misrecognition and have an impact on system usability ratings through reducing frustration and operating errors.

2 Research Hypotheses

The overall hypothesis is that, adapting the system to existing communication strategies and facilitating grounding processes should lead to improved user satisfaction. By testing systems, which use collaborative strategies known from human conversations against systems which do not, differences in usability ratings (efficiency, effectiveness, satisfaction), frustration indicators (hyperarticulation) and objective data (operation errors) are expected to be found. It is not expected that the on-demand visualization or the flexible grounding criterion affects neither gaze nor driving behavior.

In particular, the first study was set up to evaluate additional graphical visualizations, to test design alternatives and to discuss distraction potential in the dual-task context.

Transparency of the speech interaction process through state feedback should help the user avoid seeking for evidence and thus reducing the operating errors. Providing visual acknowledgment about what has been said and understood should give the user a better sense of system capabilities.

Regarding the dynamic dialogs, the purpose of the second study was to see whether the flexible grounding criterion can enhance the efficiency and effectiveness of the interaction by reducing the amount of turn-takes. Moreover, the reduction of redundant confirmation questions should reduce the operating errors and misrecognitions. Both is expected to lead to better ratings of system usability.

3 Methods

Using a deductive research approach two empirical user studies were conducted to examine the impact of visual feedback and of the flexible system grounding criterion. A control-group design was chosen in order to evaluate the effects of the innovations. While driving the simulator the probands where using the SDS for severals task concerning the adressbook (e.g. making a call, starting navigation to a contact). The tested SDS differed in visualizations (none, states, content) and dialog behavior (static, dynamic). Each usability construct was measured by at least one objective and one subjective variable. In both studies gaze behaviour and driving dynamics were recorded to analyse the impact on distraction. The system's responses were logged and the dialog behavior, such as amount of turn-takes, hyperarticulation and misrecognitions were coded by two qualified raters. In order to examine the mental models, methods like free recall and recognition tests were employed. Subjective data were collected through standardized usability questionnaires (SUMI, SUS) and interviews.

4 Solutions

The grounding process requires partners to be able to find incremental evidence of each other´s understanding. There is substantial evidence that "feedback from a spoken language system need not be in the form of speech […] it can be graphics."[5]. Accordingly, an obvious first step was to implement a system that provides a supplementary visualization of the system status and dialog results. In addition to the latter, the following system states are shown; Ready, Receiving, Processing and Speaking.
These visual indicators should spare users the effort of guessing about system status and dialog intentions [6]. Given the fact that this work is set in an in-vehicle context, the visual applications for SDS have been adjusted to demands of dual-task contexts.

The results of the first simulator study clearly indicate the positive effect of visualizations, which are showing processing information and also representing the recognized slots by the system. Furthermore it could be shown that visual feedback can trigger turn-takes and increase the perceived system transparency. It was found that an abstract presentation of the system state does not cause significant increase in the head-down time when compared to a control group without visualization. In addition to these and other findings about usability and distraction, extensive design recommendations can be stated.

Another area of action concerns the adaptation of the system behavior to the current dialog situation. By implementing a flexible grounding criterion the system will only ask for confirmation if it is insecure, similar to what humans do. In consequence, the system demands confirmation only when an increased user effort is justified. This is the case, if the previous dialog turn was difficult (low accuracy), ambient noise is present or if a misunderstanding would have serious consequences. For each dialog step, a predefined function computes, whether a confirmation request is necessary at this point of the dialog. Trimming the dialogs in this specific way can lead to efficient and satisfactory speech interaction despite the persisting shortcomings of the speech recognizers. By reducing the amount of turn-takes one automatically reduces the possibility of recognition errors. Also, it can be shown that avoiding redundant confirmations also reduces the occurrence of operating errors and hyperarticulation, which has a positive impact on system evaluation.

5 Contributions

Primarily a feedback model and a grounding criterion function for human-computer speech interaction is presented, which is based on a collaborative theory of human communication [7]. The model is used to systematically provide additional visual feedback from a SDS, thus improving and expanding the cognitive model of the user with respect to system functionality and capabilities. Additionally the grounding criterion function realizes a system that shows confirmation requests only, when necessary and reduces thereby operating errors and misrecognitions.

Results from extensive system evaluation show that it reduces both, the uncertainty of the user about the skills of the system and, in consequence, the reservations towards the voice control interface.

References

1. Yngve, V.H.: On getting a word in edgewise. In: Papers from the Sixth Regional Meeting of Chicago Linguistic Society, pp. 567–578. Chicago Linguistics Society, Chicago (1970)
2. Burnett, G.E., Joyner, S.M.: An assessment of moving map and symbol-based route guidance systems. In: Ian Noy, Y. (ed.) Ergonomics and Safety of Intelligent Driver Interfaces, pp. 115–136. Lawrence Erlbaum Associates, Mahwah (1997)
3. Grice, H.P.: Logic and conversation. In: Cole, P., Morgan, J.L. (eds.) Syntax and Semantics III: Speech Acts, vol. 58, p. 41. Academic Press, New York (1975)
4. Clark, H.H., Wilkes-Gibbs, D.: Referring as a collaborative process. Cognition 22, 1–39 (1986)
5. Brennan, S.E., Hulteen, E.: Interaction and feedback in a spoken language system: A theoretical framework. Knowledge-Based Systems 8, 143–151 (1995)
6. Brennan, S.E.: The grounding problem in conversation with and through com-puters. In: Fussell, S.R., Kreuz, R.J. (eds.) Social and Cognitive Psychological Approaches to Interpersonal Communication, pp. 201–225. Lawrence Erlbaum, Hillsdale (1998)
7. Clark, H.H., Schaefer, E.F.: Contributing to discourse. Cognitive Science 13, 259–294 (1989)

Development of a Methodology for Evaluating the Quality in Use of Web 2.0 Applications*

Tihomir Orehovački

University of Zagreb, Faculty of Organization and Informatics,
Pavlinska 2, 42000 Varaždin, Croatia
tihomir.orehovacki@foi.hr

Abstract. Quality in use is comprised of two seemingly different though interlocking concepts: usability and user experience. Consequently, complementary evaluation of pragmatic and hedonic attributes could significantly affect the acceptance of software applications. However, in the context of Web 2.0 applications this topic has still not attracted enough attention from the HCI community. Therefore we present a research aimed at developing a methodology that would facilitate the analysis and comparison of evaluated Web 2.0 applications.

Keywords: Web 2.0, Quality in Use, Usability, User Experience, Subjective and Objective Measures, Evaluation Methodology.

1 Introduction

Quality in use is considered to be one of the most important factors that affect a wide acceptance of software applications. It is comprised of two seemingly different but rather complementary concepts [6]. Usability refers to the product-centred evaluation of pragmatic attributes through the use of both subjective and objective measuring instruments. On the other hand, user experience (UX) consists of applying subjective measuring instruments on the assessment of hedonic attributes [5]. HCI literature contains a number of different models, methods, and standards aimed at evaluating the quality and usability of software applications like information systems, web sites, web portals and alike. However, the existing research related to the evaluation of Web 2.0 applications in general is fairly scarce. We strongly believe that this is accounted for by the inadequacy of current approaches for their evaluation, as confirmed by several recent studies. According to the research in Hart et al. [4], the popular social networking site Facebook should not be an example of a success story because it complies with only two of ten heuristics originally proposed by Nielsen [7]. Hart et al. also reported that attributes such as ease of use, usefulness and playfulness have a major impact on user behavioural intentions. A study into usability evaluation [11] revealed that YouTube had also met only two heuristics. Thompson and Kemp [14] argued that the emphasis on user experience is one of the main reasons why Web 2.0

* Supervisors: Prof. Dragutin Kermek, Ph.D. (dragutin.kermek@foi.hr) and Prof. Andrina Granić, Ph.D. (andrina.granic@pmfst.hr).

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 382–385, 2011.

applications such as Flickr, Wikipedia and YouTube have a large number of active users. Moreover, they extended the Nielsen's set of traditional heuristics with two additional heuristics although the validity of that modified set has not been empirically confirmed. The findings stated above motivated us to initiate a research on the development of an evaluation methodology that would extend the concept of quality in use in the context of Web 2.0 applications.

2 Research Questions

In line with the above, we outline the following research questions:

a. What are the most important subjective and objective quality attributes that contribute to users' satisfaction and loyal behaviour?
b. How can these quality attributes be measured? In particular, are the existing measuring instruments suitable for evaluating the quality in use of Web 2.0 applications?
c. Can these quality attributes be translated into a set of heuristics that would facilitate the assessment of Web 2.0 applications?
d. Can these quality attributes be converted into a set of guidelines that would contribute to improved design of Web 2.0 applications and their testing?
e. Which methods, techniques and procedures should constitute an evaluation methodology that simplifies an analysis and comparison of Web 2.0 applications?

3 Methodology

Our research is divided into three parts: theoretical background, pilot study and main research. Until now, the theoretical part and the pilot study have been completed.

In the theoretical part, a literature review was prepared according to the methodology proposed by Webster and Watson [15]. An analysis of recent relevant research in the field of web quality and usability assessment resulted in a set of attributes that may have a significant role in the evaluation of the quality in use of Web 2.0 applications [10]. Based on their theoretical similarities, the attributes were initially placed into six different categories: system quality, service quality, information quality, performance, effort, and acceptability. In addition, based on the type of the measuring instrument to be used for data collection (a questionnaire or a tool for logging actual use), the attributes were grouped into subjective and objective quality attributes. The validity and reliability of questionnaire items was evaluated according to the guidelines suggested by Straub et al. [12].

The aim of the pilot study was to test the suitability of the measuring instruments for evaluating the quality in use of Web 2.0 applications. The participants in the pilot study were students who used Web 2.0 applications to perform various educational activities. In the first experiment, two methods were employed in order to gather both quantitative (questionnaire) and qualitative (retrospective thinking aloud, RTA) data about the perceived quality of cloud based applications for collaborative writing [9]. In the second experiment [8], the research methodology was extended with the logging actual use method. The subjective assessment of Web 2.0 applications for mind mapping was conducted with the use of a questionnaire and RTA while the objective data about mouse movements, mouse clicks, keystrokes and time on task were gathered by means of the Mousotron tool [1]. An analysis of the data obtained from both experiments revealed a list of matching attributes that influence users' satisfaction and loyal behaviour regardless of the type of application used, including ease of use, effectiveness, reliability, interactivity, customizability, and navigability.

The remaining part of the research is structured as follows. Firstly, two different groups of experts (HCI researchers and web developers) will categorize quality attributes using the Q-sorting method [13]. Secondly, a scenario-based evaluation of Web 2.0 applications with different functionalities will be conducted. Data will be collected using a questionnaire (perceived quality in use) and a tool for logging actual use (estimated quality in use). Participants in the study will be recruited from different departments of two Croatian universities. The data obtained from the experts will be used in the development of a conceptual model that will be composed of two sub-models: a measurement and a structural model, respectively. The former will take into account relationships between each latent variable (category) and corresponding manifest variables (attributes), while the latter will take into account the relationships among the latent variables. The validation of the conceptual model will be carried out with Partial Least Squares (PLS) path analysis in accordance with the catalogue of criteria discussed in [3]. Measurement models will be assessed by means of content validity, indicator reliability, construct reliability, convergent validity and discriminant validity. On the other hand, the structural model will be evaluated in terms of its effect size and predictive validity. Based on the outcomes of the Q-sorting and the PLS analysis, a tree with logical relationships between categories and attributes will be specified and elementary criteria for each attribute will be defined. The final estimate (score) of quality in use will be calculated using a stepwise process of logic aggregation of preferences [2].

4 Expected Contributions

The scientific contribution of the research is two-fold. From a theoretical perspective, the knowledge and experience in the field of quality and usability evaluation of Web 2.0 applications will be systematized. The most important quality attributes that significantly affect users' satisfaction and loyalty will be identified. Moreover, it will be determined whether the weight of each attribute varies depending on the Web 2.0 application type and the context of use.

From a practical point of view, the new methodology will offer a number of benefits. Firstly, it is expected to enable the evaluation of Web 2.0 applications using both subjective and objective measuring instruments, which may alleviate possible shortcomings of their separate use. Secondly, through the combination of several data collection methods, the process of detection of problems in Web 2.0 application usage will be facilitated and accelerated. It is noteworthy that, apart from the HCI experts and users, web developers (programmers, designers, architects) will also take part in the research (which so far, to our knowledge, has been rather rare). Thirdly, based on the results obtained from the practical use of the methodology, a new set of heuristics aimed for the evaluation of Web 2.0 applications will be created. In addition, the research will result in a series of guidelines and recommendations for the development of Web 2.0 applications. Finally, the afore-described methodology will enable the "transformation" of a lot of quantitative data into a single score, which would facilitate the analysis and comparison of evaluated Web 2.0 applications.

References

1. Blacksun Software (2011), http://www.blacksunsoftware.com/mousotron.html
2. Dujmović, J.J.: Preferential Neural Networks. In: Antognetti, P., Milutinović, V. (eds.) Neural Networks – Concepts, Applications, and Implementations, pp. 155–206. Prentice Hall, New Jersey (1991)
3. Esposito, V.V., Chin, W.W., Henseler, J., Wang, H.: Handbook of Partial Least Squares: Concepts, methods, and applications. Springer, Berlin (2009)
4. Hart, J., Ridley, C., Taher, F., Sas, C., Dix, A.: Exploring the Facebook Experience: A New Approach to Usability. In: 5th Nordic Conference on Human-Computer Interaction: Building Bridges, pp. 471–474. ACM, Lund (2008)
5. Hassenzahl, M., Tractinsky, N.: User experience – a research agenda. Behaviour & Information Technology 25(2), 91–97 (2006)
6. ISO/IEC 25010:2011. Systems and software engineering – Systems and software Quality Requirements and Evaluation (SQuaRE) – System and software quality models (2011)
7. Nielsen, J.: Heuristic evaluation. In: Nielsen, J., Mack, R.L. (eds.) Usability Inspection Methods, John Wiley & Sons, New York (1994)
8. Orehovački, T., Granić, A., Kermek, D.: Exploring the Quality in Use of Web 2.0 Applications: The Case of Mind Mapping Services. Lecture Notes in Computer Science (accepted for publishing)
9. Orehovački, T.: Perceived Quality of Cloud Based Applications for Collaborative Writing. In: Pokorny, J., et al. (eds.) Information Systems Development – Business Systems and Services: Modeling and Development. Springer, Heidelberg (in press)
10. Orehovački, T.: Proposal for a Set of Quality Attributes Relevant for Web 2.0 Application Success. In: 32nd International Conference on Information Technology Interfaces, pp. 319–326. IEEE Press, Cavtat (2010)
11. Silva, P.A., Dix, A.: Usability – Not as we know it! In: 21st British HCI Group Annual Conference on HCI 2007: People and Computers XXI: HCI...but not as we know it, vol. 2, pp. 103–106. ACM, University of Lancaster (2007)
12. Straub, D., Boudreau, M., Gefen, D.: Validation Guidelines for IS Positivist Research. Communications of the Association for Information Systems 13(1), 380–427 (2004)
13. Thomas, D.M., Watson, R.T.: Q-Sorting and MIS Research: A Primer. Communications of the Association for Information Systems 8, 141–156 (2002)
14. Thompson, A.-J., Kemp, E.A.: Web 2.0: extending the framework for heuristic evaluation. In: 10th International Conference NZ Chapter of the ACM's Special Interest Group on Human-Computer Interaction, pp. 29–36. ACM, New Zealand (2009)
15. Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: writing a literature review. MIS Quarterly 26(2), xiii–xxiii (2002)

Distortion Techniques for Sketching Interaction

Paul Schmieder*

Department of Computer Science,
University of Auckland,
Private Bag 92019, Auckland, New Zealand
psch068@aucklanduni.ac.nz,
{beryl,john}@cs.auckland.ac.nz

Abstract. Using sketching as the application domain, this research compares current distortion techniques for enlarging content. The goal is to develop a distortion lens which allows for a natural and uncomplicated drawing and writing experience on an electronic device.

Research Area: Distortion techniques to enlarge interaction spaces for ink.

1 Introduction and Motivation

The problem of limited display size on devices such as mobiles and laptops is well known and studied. Techniques that allow users to get a better overview or see more details such as pan and zoom (commonly referred to as overview and detail techniques) are very popular and widely applied [1], but are unsuitable for tasks where the complete content has to be visible at all times. By using distortion effects, focus and context techniques don't occlude content while showing zoomed parts in more detail [2].

The main problem when enlarging one area of a display is the additional space required for the enlarged part – perhaps occluding other parts. To avoid occlusion, focus and context techniques distort the content to generate the required space. For example, fisheye lenses [3, 4] use a central focus region for the detailed view with a distorted transition region surrounding it. The transition region generates the additional space required while providing a visual transition between the focus and context regions. Another distortion technique is sigma lenses [5] which use translucency to create a transition between the focus and context region. Sigma lenses do not require a separate transition region as the transition happens in the focus region thus saving valuable canvas space.

Distortion techniques have been the subject of studies for some time and have been proven useful, despite their disadvantages, in various domains such as viewing maps, selecting targets, and steering tasks [2, 7]. However, these distortion techniques have been evaluated for passive tasks, where content is viewed and selected, but not for active tasks of creation and editing. For the remainder of this paper the term 'active use' refers to creating and editing content and the term 'passive use' refers to viewing and selecting content.

* This dissertation work is supervised by Beryl Plimmer and John Hosking.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 386–389, 2011.

The practicality and effectiveness of applying zooming techniques for active use for sketching has been shown by Agrawala and Shilman [8]. In their interface a region of the document is magnified to allow annotations to be written at the same size as the document's content. Their evaluation showed that zooming was useful when editing and writing in areas where space was limited, but participants found it less useful when sketching long gestures such as underlines and circles.

While Agrawala and Shilman [8] used a zoom region which covered parts of the content, Lank and Phan [9] employed a fisheye lens to avoid hiding context. Due to limited hardware performance Lank and Phan's lens remained at the position where it was activated rather than following the pen. Their use of a static lens helped them avoid the problems faced by movable lenses. Although their work is a first step toward the successful utilization of distortion techniques for active use, the static lens was found to be heavily constraining. It was identified in their evaluation study as the prototype's major drawback by a number of the participants [9]. Yet, the lens was, overall, perceived positively and despite this key limitation participants preferred the fisheye lens over having to scroll the canvas for their drawing tasks.

In summary, previous work has shown the usefulness of distortion techniques to tackle problems of limited space for passive use. Additionally, it indicates the potential for supporting interaction where content is created and edited although the realization is somewhat limited. These findings, in addition to my own experience in sketching[10, 11], led me to believe that new/novel distortion techniques can be of great benefit to better utilize limited canvas space for *active use.*

2 Working Hypothesis

Distortion techniques have potential for active use. In particular, the active use of distortion techniques is perceived positively by potential users and it results in performance gains. Additionally, the increased mental workload caused by distortion techniques does not hamper the user's creation and editing activities. To achieve this, current distortion techniques such as [3, 5-9] may have to be adjusted or a new technique developed such that their shortcomings for active use are resolved.

3 Methodology and Conducted Research

In this research we use digital sketching to investigate distortion techniques and their impact on the creation and editing of content. Sketching is an excellent application area as it provides users with an intuitive way to express and explore ideas. When designing a lens for more than just viewing and selecting content, there are many different aspects to consider. As distortion already puts additional mental workload onto the user we want to avoid using techniques that add to that workload. For example, we want to keep the distortion applied to a stroke being drawn to a minimum, especially the parts of the stroke close to the pen tip. For this research a sketch tool prototype has been implemented including different distortion lenses and several overview and detail techniques.

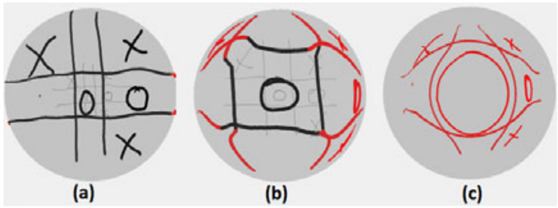


Fig. 1. Three different lenses showing a noughts and crosses game. Lens (a) is a magnifier. Lens (b) and (c) use a distorted transition region to generate the space required by the enlarged region. Lens (b) has a flat enlarged region while lens (c) has only a point at maximum magnification. Lenses (a) and (b) show the undistorted ink using translucency.

To determine which distortion techniques are suitable for active use and whether the current techniques have to be optimized or completely reworked, there are several steps to be taken. First, the task categories involved for active use have to be investigated to see whether each category requires a different lens; previous research indicates that different lenses are preferred by users when drawing and writing [8]. To study the different task categories, an evaluation study will be conducted where the participants are given drawing and writing tasks.

Second, distortion lenses and their configurations have to be developed for drawing and writing. Lenses can use different distortion techniques such as fisheye and sigma lenses. Within each technique there are a variety of options to choose from, such as the focus region being a single point or a flat area. The distortion function influencing the transition between the focus and context regions is another configuration decision. Once determined, many of these options can be fixed, though some options such as zoom level and size must be dynamically adjustable by the user. As there are complex relationships between the different options, a series of evaluation studies is necessary to answer these questions. Ultimately, the analysis of current distortion techniques may result in existing techniques being revised for active use or requiring a novel technique.

Third, having identified and implemented the preferred lenses, the lenses need to be compared to non-distortion techniques such as pan & zoom and radar. The results will answer the main research question whether and which distortion techniques are suitable for active use. Therefore, a comparative study will be undertaken examining the performance of the different techniques in detail. Fourth, once the appropriateness of distortion techniques is confirmed, the lenses and their interplay have to be investigated and optimized. This includes the transitions between the lenses (e.g. when the user switches between drawing and writing) as well as how the lens is best operated (e.g. target acquisition) to optimize usage.

Since there is a considerable amount of work until we can answer our main research question we have conducted an informal study to see whether distortion techniques, even if not optimized, can compete with other techniques. We have asked students in our research group to complete a series of writing and drawing tasks using one lens as well as the radar and pan & zoom techniques. Overall, they preferred using the fisheye technique but mentioned that the lens was less suitable for writing because it was too small. When parts of the word reach the transition region, these parts are distorted before one finishes writing the word, which users found confusing. These findings are similar to Lank's evaluation [9].

4 Proposed Solution

The lens will be designed according to the requirements described in the previous section. Specifically, the lens we propose will enlarge a part of the canvas while providing a transition between focus and context regions so that the complete content is visible at all times. Therefore, the lens may use newly developed distortion techniques or an optimized combination of already existing techniques. The lens will be optimized for active use and employ effective transitions between modes.

5 Expected Research Contributions

Our main research contribution will be to show the suitability or unsuitability of distortion techniques for active use. This includes known distortion techniques as well as novel techniques that may be developed in this research. We hope to show that the additional mental effort to comprehend the effect of distortion does not have a negative impact. Furthermore, we believe the methods used to transition between the distortion lenses, as well as the mode changes to control zoom and lens size will be applicable in other domains. Ultimately, the goal is that our lens toolkit will allow users to create and edit content faster and more conveniently than traditional overview and detail techniques on canvases where interaction space is limited.

References

1. Plaisant, C., Carr, D., Shneiderman, B.: Image-browser taxonomy and guidelines for designers. IEEE Software 12(2), 21–32 (2002)
2. Leung, Y., Apperley, M.: A Review and Taxonomy of Distortion-Oriented Presentation Techniques. ACM Transactions on Computer-Human Interaction (TOCHI) 1(2), 126–160 (1994)
3. Furnas, G.: Generalized fisheye views. ACM SIGCHI Bulletin 17(4), 16–23 (1986)
4. Spence, R., Apperley, M.: Data base navigation: an office environment for the professional. Behaviour & Information Technology 1(1), 43–54 (1982)
5. Pietriga, E., Appert, C.: Sigma lenses: focus-context transitions combining space, time and translucence. In: Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems, pp. 1343–1352. ACM, Florence (2008)
6. Gutwin, C.: Improving focus targeting in interactive fisheye views. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing our World, Changing Ourselves, pp. 267–274. ACM, Minneapolis (2002)
7. Carpendale, M., Montagnese, C.: A framework for unifying presentation space. ACM, New York (2001)
8. Agrawala, M., Shilman, M.: DIZI: A Digital Ink Zooming Interface for Document Annotation. In: Human-Computer Interaction, INTERACT 2005, pp. 69–79 (2005)
9. Lank, E., Phan, S.: Focus+Context sketching on a pocket PC. In: CHI 2004 Extended Abstracts on Human Factors in Computing Systems, pp. 1275–1278. ACM, Vienna (2004)
10. Schmieder, P.: Comparing Basic Shape Classifiers: A Platform for Evaluation Sketch Recognition Algorithms. In: Computer Science, p. 190. University of Auckland, Auckland (2009)
11. Schmieder, P., Plimmer, B., Vanderdonckt, J.: Generating Systems from Multiple Sketched Models. Journal of Visual Languages & Computing (2009)

Evaluation of Information Classification on Websites and Impact of Culture: A Cross Country Comparison of Information Classification

Ather Nawaz

Department of Informatics, Copenhagen Business School,
Howitzvej 60, DK-2000, Copenhagen, Denmark
an.inf@cbs.dk

Abstract. The structure of information classification has an important role in the usability of websites. A majority of cross cultural studies have emphasized on localized elements of interface design and termed them as cultural markers However, not many studies have pointed out on how the classification of information on the websites can be similar or different, especially for those communities who have recently joined the global community of the Internet. This research aims to investigate the information classification of users in different countries and compares it with information classification of a website. The study will evaluate how different demographical properties impacts on the information structure of websites. The study will also evaluate to what extent users' performance and productivity changes, when the information classification of website matches with the end users mental model.

Basic information of PhD research

PhD Supervisor: Torkil Clemmensen, Department of Informatics, Copenhagen Business School, Denmark. Email: tc.inf@cbs.dk

The Research area of my work: The research area of my work is usability and sub-area of my work is website Information classification using card sort in cultural context.

Brief description of research Topic: This research investigates the information classification of websites from a cultural perspective and compares the website information classifications with the users' mental model.

1 Introduction

Due to the abundance of information, it has become a challenging problem to track and retrieve information from institute, university or e-commerce websites. The challenge may however not be the same in different countries. One of the key issues in the design of websites is the information classification on websites. With the global penetration of the internet, especially into regions where computer use has hitherto been scarce, the diverse needs and perceptions of quality place a special demands

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 390–393, 2011.

upon the web page design and how information is placed [1]. The information organized in the correct location speeds up the information retrieval for the users and provides a good user experience of the website. In the cross cultural studies of websites, much of the work has focused on similarities and differences between Asia and the West and little work appears to have been done to investigate the usability, language bias, and structure of websites in those communities who have recently joined the global community of the Internet [2].

Our research adds the body of knowledge in the domain of information structure on the website and users' understanding with the information structure on the websites especially in those countries where there is not much research done. The research proposes how the information classification in the different countries has the properties that do not resemble to each other, despite the similarity of participants' years of study and age.

1.1 Research Problem

With the access of large amount of information on corporates, institutes and e-commerce websites, it is often a challenge to retrieve information on these websites. The challenge may, however, not be the same in different countries. A key issue in a good website design is the classification of the information on the website. If the website information is classified in a manner that fits well with the user's perception of the topic, then information retrieval on the website is efficient and may even be satisfying. Most of the cross-cultural studies of websites have focused on the usability, language biases, and structure of Asian and Western websites. Little work appears to have been done investigating the structure of the websites in communities that have recently joined the global Internet community. Furthermore it is not the technical aspects that are cumbersome but the mental model of the end users in a group that can be different from the information classification of the websites. In order to improve the information findability and performance of users, demographic diversity and its influence on information classification for the websites should be acknowledged.

The prior research addresses the issue by emphasizing on the interface design aspect of websites and cultural markers but did not stress on the classification criteria of contents. A few of the researches talk about the classification models on websites. The studies of Choong [3] shows how the users from two different locations have different mental models and conceptions about the classification of information. The studies address the cultural aspect of information classification but only highlight the Chinese and American perceptive on information classification.

1.2 Research Question

1. How is the perception of website structure similar and/or different in the countries that have recently joined the global community of internet?
2. How does the fit between the users' mental model and the website's information classification affects the *information findability* on the websites?
3. To what extent does the information literacy of technology interaction influences the perception of information classification on websites.

2 Method

The research for my PhD is conducted in two stages. In the first stage, I collected the data from the users of a university website in Pakistan and Denmark. The first stage of the research helped me to identify the important and interesting factors of the research. I used brainstorming and card sorting technique to understand the users' mental model. Card sorting is a technique aligned with the Kelly's personal construct theory [4]. It assumes that people make sense of the world by classification and people can describe their own classification of the world with reasonable validity and reliability [4, 5]. Card sorting is a useful way of involving users with the classification of information into groups. It provides an insight into users' mental models, illuminating the way they often tacitly group, sort and label tasks and content within their own heads[5]. The supplementary activities included information findability activities and surveying the users' satisfactions in each group. The 1st stage of the research facilitated in understanding the factors related to the website structure and understanding the users' mental models.

The results of 1st study showed interesting results from the Pakistani groups of users. The results of the study showed how the end users perceive university websites bilingually in a different way. Two websites were used for information findability. The brainstorm card sort session also reviled interesting results. The primary results provided the culturally specific preferences to with the university website in both countries. For the Pakistani group, the participants of the 1st study intended for bilingual website with bilingual contents presentation on the same webpage which was not present in their university website. The result of the study also showed [illegible]

Study 2: In the second stage of the research, the card sorting activities will be revised with learning from study 1 of first stage of data collection but this time with the same website across groups. The experiments on a website for home appliance will be used to manipulate the information structure of these websites in Denmark, Pakistan and Malaysia.

2.1 Theoretical Lenses for the Research

In many studies, it is argued that the collective or interdependent nature of Asian society takes the contextual view of the world and believes that events are highly complex and determined by many factors such as thought process, language and cultural background [6]. On the other hand, the individualistic or independent nature of western society seems consistent with the western focus on particular objects in isolation from their context and with westerns' belief that they can know the rules governing objects and therefore can control the objects' behaviour. The difference in their approach of has an influence on the resulting product. Martine & Rugg [7] measured the perceived similarity of webpages using a card sort technique. The card sorting method appears particularly well suited for a comparative and knowledge representation oriented study of information classification on a website.

3 A Sketch of Proposed Solution

From my PhD research, I intend to propose a list of factors that are involved in the information classification of websites. The study will list the factors that affect the performance of users and how they can be minimized to increase performance. The outcome of the study will provide the suggestion to avoid the bias in the interface design and information structure of the websites that is brought by the different understanding of users towards websites. This research will be helpful for software designers and web developers who are designing and developing websites and applications for Pakistan, Malaysia and Denmark. It will also be helpful for the companies who are localizing their applications in other parts of the world.

References

1. Aykin, N.: Usability and internationalization of information technology. Lawrence Erlbaum Associates, Mahwah (2005)
2. El Said, G., Hone, K.: Cross-Cultural Web Usability: An Exploration of the Experiences of Egyptians Users. In: International Workshop on Internationalisation of Products and Systems (IWIPS 2001), Milton Keynes (2001)
3. Choong, Y.-Y.: Design of computer interfaces for the chienese population, p. 180. Purude University (1996)
4. Kelly, G.: The psychology of personal constructs. Clinical Diagnosis and Psychotherapy 2 (1991)
5. Rugg, G., McGeorge, P.: The sorting techniques: a tutorial paper on card sorts, picture sorts and item sorts. Expert Systems 14(2), 80–93 (1997)
6. Nisbett, R.E.: The geography of thought: Why we think the way we do. Free Press, New York (2003)
7. Martine, G., Rugg, G.: That site looks 88.46% familiar: quantifying similarity of Web page design. Expert Systems 22(3), 115–120 (2005)

Exploring New Ways of Utilizing Automated Clustering and Machine Learning Techniques in Information Visualization

Johann Schrammel

University of Salzburg, 5020 Salzburg
schrammel@cure.at
Advisor: Prof. Manfred Tscheligi
manfred.tscheligi@sbg.ac.at

Research Area: Information visualization, human-computer interaction.

Research Topic. The main research topic of the thesis is to explore the possibilities of automated clustering and machine learning techniques for developing new approaches in information visualization.

Research Problem. The main goal of information visualization is to present data to the users in a way that optimizes intelligibility of the data and support the detection of relevant patterns in the data, where the application context defines what qualifies as 'relevant'. Many different approaches typically tailored to a specific problem have been developed within the past years. At the same time the application of mathematical methods for data analysis and identification of patterns has substantially increased, and is typically referred to as data mining. Different visualization techniques are used in data mining, however the systematic and dynamic integration of data mining techniques with visualization approaches is only in its beginning.

The main research problem tackled in the thesis is how to best integrate approaches from traditional data mining and information visualization domains to achieve more usable and helpful analysis tools for large amounts of data. The basic visualization approach explored is tag clouds, and especially concepts in displaying them according to a semantic structure as proposed e.g. by [2,3,1] are further developed. The overall research problem can be divided into sub-problems which need to be addressed in order to develop actually useful approaches:

- Develop new approaches and concepts for integrating data mining and information visualization techniques
- Develop a detailed understanding of the human perception and analysis process involved when interacting with these systems
- Develop a understanding on the impact of errors, artifacts and imperfection in the data analysis on experienced quality and usefulness
- Define quality benchmarks that are needed to be met to provide real value and benefit to the user when using such systems
- Evaluate the developed concepts and prototypes in realistic scenarios, to be able to estimate their value and applicability in real-world contexts.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 394–397, 2011.

Relevance of research. We think research on this topic is highly important. The information society is producing more and more data, and better means to make sense of the available data are urgently needed. Also, new visualization approaches can help to communicate complex affairs in an intelligible manner, which is relevant in ever more context. Examples that illustrate this increasing need to understand complex data even in everyday situations are e.g. home control units in smart grid applications or interfaces for the specification of security, firewall and privacy settings.

The research hypothesis (claim). The overall research hypothesis is that utilizing automatically calculated relations between data elements and visualizing these identified relations in a proper way can help to develop a deeper understanding of the data. We also expect such approaches to support the user in performing different types of typical tasks such as data exploration or specific searches more efficiently.

Within the thesis work I plan to research this overall hypothesis and the related sub-aspects in a series of prototypes which explore the possibilities and potentials of different approaches. As of now the following design approaches already have been explored (a and b) respectively are planned (c to f):

a) Semantic layout of data elements: data elements are placed according to a calculated semantic layout based on the similarity of elements
b) Bottom-up clustering: Data elements are clustered automatically, and no label for the identified clusters is available.
c) Top-town categorization: Categories are defined a priori, and learned using standard machine learning algorithms using manually labeled training data.
d) Differential placing of elements: data elements are placed according to their similarity with a predefined bipolar concept.
e) Non-exclusive grouping: Data elements are cluster automatically into groups, however in contrast to b) the grouping is non-exclusive i.e. the categories do overlap.
f) Time-series visualization: The development of data structures across time is visualized

We hope to identify further promising approaches as part of the discussion in the doctoral consortium.

Methods used. The research approach follows a standard process consisting of the design of new approaches, a prototypical implementation of these approaches, optimization of these prototype implementations through fine-tuning of parameters based on informal user tests and a formal empirical evaluation with real end users in a realistic application context.

Empirical evaluation will target three main areas: performance, subjective satisfaction and analysis of the perception process. Performance will be operationalized as task completion time and error rate, subjective satisfaction with the approach will be explicitly asked in a questionnaire and analysis of perception process will be done using eye tracking technology. The main evaluation methodology will be similar to the approaches applied in the already performed evaluations as reported in [5] and [7].

A sketch of the proposed solution. In detail the following test prototypes have been implemented respectively are planned as of now:

a) Semantic placing of elements [5]: A thirst approach was to use semantic placing of data elements in a tag cloud. We used the getrelated-function of flickrs API to retrieve a list of the tags most related to each word within the tag cloud. Then based on the number of co-occurring related tags a measure for the relatedness of two tags was calculated. An alternating least-squares algorithm to perform multidimensional scaling (ALSCAL) was used to compute a two-dimensional arrangement of the tags. In the third step we used the value on the y-axis to form 7 groups of 11 resp. 10 tags each. Next tags within each group were sorted according to their value on the x-axis. The result provided an 11 times 7 arrangement that was used to generate the tag cloud.

b) Clustered tag clouds (related submission #232): Here we calculate tag similarity using a well proven method known as Jaccard coefficient. Similarity between tags is measured by the intersection divided by the union of the sample set. Based on this similarity measures clusters of tags where calculated using the bisecting k-means approach. For a discussion of different clustering approaches and their pros and cons see Steinbach et al. [8]. The clusters were calculated using the CLUTO-Toolkit provided and described by [4]. Basically the N-dimensional similarity matrix of tags was used as an input for the clustering algorithm. The target number of clusters to calculate was specified as 20. This number was chosen to form clusters of about 5 tags, which informal pre-test showed to be a good size for clusters.

c) Categorized tag clouds (under construction): The clustering process used in b) doesn't identify the topic of a cluster but only groups similar items together. To be able to also study the effects of labeled groups we plan to apply machine learning algorithms for the classification process, in which a subset of the data is categorized and labeled manually and used as training data.

d) Differential tag clouds (planned): Within this approach the idea is to use the calculated similarity to predefined bipolar concepts (e.g. work versus leisure) to arrange the items on the screen. Similarity is expected to be calculated by the same method as in c), however more extensive training data on the bipolar concepts are needed. We also plan to use interactivity in this approach, i.e. to allow the user to select different concepts and to directly adapt the display.

e) Non-exclusive grouping (planned): In this approach we plan to apply non-exclusive clustering algorithms to the data set and to develop different visualizations for the users. A possible approach is to implement a slider, which dynamically changes the threshold above which the groups are displayed, or to show the related items only on mouseover to minimize visual clutter. We also plan to explore the possibilities of automatic drawing approaches for Euler graphs [5].

f) Time-series visualization (planned): Here we plan to introduce a dynamic element. The basic idea is to use the developed automated classification and placement algorithms to visualize the development across time within a field. For example, when using a differential tag clouds approach (compare item d) with the concept work – leisure and items are arranged horizontally regarding their distance to the concept (e.g. work-related displayed left, leisure related displayed right) and to also indicate the allocation to one of these concepts using color. One then could analyze chronological data sets (e.g. articles in a newspaper per day) and show a dynamic display of the development of the topics over time.

The expected contributions of the PhD research

a) New methods for displaying data that combine data mining and information visualization approaches that allow developing a faster understanding of patterns within the data and support users in complex analysis tasks.
b) Improved understanding of perception processes in information visualization especially with regard to the perception of modified tag clouds

References

1. Fujimura, K., Fujimura, S., Matsubayashi, T., Yamada, T., Okuda, H.: Topigraphy: visualization for large-scale tag clouds. In: Proc. WWW 2008 (2008)
2. Hassan-Montero, Y., Herrero-Solana, V.: Improving tagclouds as visual information retrieval interfaces. In: Proc. InfoSciT 2006 (2006)
3. Lohmann, S., Ziegler, J., Tetzlaff, L.: Comparison of Tag Cloud Layouts: Task-Related Performance and Visual Exploration. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5726, pp. 392–404. Springer, Heidelberg (2009)
4. Karypis, G.: CLUTO - a clustering toolkit. Technical Report #02-017 (November 2003)
5. Rodgers, P., Mutton, P., Flower, J.: Dynamic Euler Diagram Drawing. In: Visual Languages and Human Centric Computing (2004)
6. Schrammel, J., Leitner, M., Tscheligi, M.: Semantically structured tag clouds: an empirical evaluation of clustered presentation approaches. In: Proc. CHI 2009 (2009)
7. Schrammel, J., Deutsch, S., Tscheligi, M.: The Visual Perception of Tag Clouds - Results from an Eye Tracking Study. In: Proc. INTERACT 2009 (2009)
8. Steinbach, M., Karypis, G., Kumar, V.: A comparison of document clustering techniques. In: Grobelnik, M., Mladenic, D., Milic-Frayling, N. (eds.) KDD 2000 Workshop on Text Mining, Boston, MA (2000)

Exploring Serendipity's Precipitating Conditions

Lori McCay-Peet

Dalhousie University, Faculties of Management and Computer Science,
6100 University Avenue, Halifax, NS, Canada
mccay@dal.ca

Abstract. Serendipity is generally characterized as a sagacious, unsought discovery. Innovations and advances in science and engineering such as penicillin and Teflon as a result of interactions with tangible engineered and natural phenomena are often labeled serendipitous. Serendipity also results from more conceptual interactions with information, knowledge, and ideas. But how does serendipity emerge when the discovery is predominantly conceptual? What conditions in the digital environment would help facilitate less tangible but similarly unexpected and fortunate interactions for knowledge workers? The objective of my research is to approach the study of serendipity as a process involving precipitating conditions, interacting internal and external factors that either hinder or facilitate serendipity, to understand how to best facilitate serendipity in a digital, information-rich environment. This research will contribute to an emerging field of study – support for serendipity in information systems – that is striving to make our experiences in digital environments richer and more meaningful.

Keywords: Serendipity, precipitating conditions, information systems.

1 Introduction

Serendipity is an unexpected experience characterized by an anomalous observation and valuable outcome and dependent on an individual's strategic insight [1]. Primarily associated with scientific and engineering discoveries and developments, serendipity also results from more conceptual interactions. Serendipity, in management, for example, sparks ideas and new opportunities that drive strategic decisions [2]. In education, serendipity facilitates new and interesting entries to educational material and learning [3]. Despite research indicating serendipity's significance across a number of fields [4] and the pervasive and persuasive anecdotal documentation of its importance [5], information-rich environments such as digital libraries primarily support targeted search interactions. Absent are features that are intended to guide users to interactions with divergent though potentially useful content. As well, research to date has primarily focused on the need to deliver unexpected, surprising, and novel content to the user, often neglecting other precipitating conditions – factors, for example, influencing the receptivity of users to this potentially serendipitous content such as affect and motivation [6] and affordances in the environment [7].

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 398–401, 2011.

2 Research Problem and Hypothesis

The process of serendipity in knowledge work (Figure 1) illustrates the movement of an individual through the stages of serendipity. This model, adapted from Cunha's [2] model of the process of serendipity in organizational management has five main elements and unfolds as follows: 1) an individual experiences an information encounter [10] or makes an anomalous observation [1] (the trigger) 2) and an association between the object and the individual's knowledge and experience or a bisociation between ideas is made. The individual 3) follows-up on the association/bisociation 4) and a valued outcome is the potential result. This process, however, is influenced by the fifth element of the process: the convergence of precipitating conditions. Precipitating conditions influence each stage of the process, impacting, for example, whether associations/bisociations and follow-up are dismissed or delayed, leading to an incubation period [8] or *negative serendipity* [9] in which the opportunity for a valuable outcome is taken by someone else.

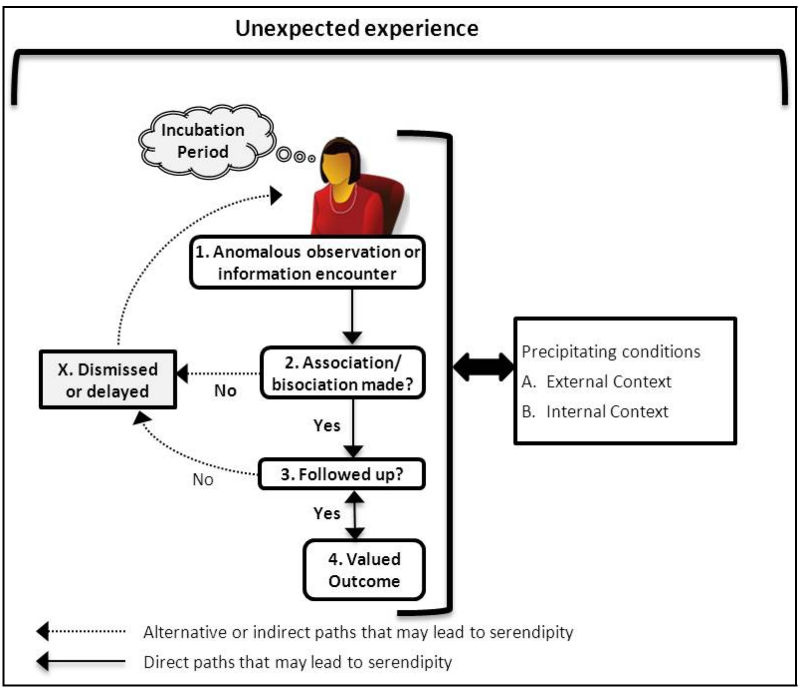


Fig. 1. Serendipity in knowledge work (adapted from [2], [8])

Precipitating conditions of serendipity may have a positive or negative impact on the process. They are conditions that influence an individual's ability and inclination to become aware of, respond to, and follow up on a potentially serendipitous experience, and ultimately impact the outcome. Precipitating conditions include A) external and B) internal context. External context may include, for example, the design of an information space such as a library [7], other people who may help spark a bisociation [8], or an imposed task that leads an individual to ignore useful but not immediately relevant information [10]. Internal context refers to the actions and characteristics of the individual. Internal context includes, for example, misleading preconceptions that may impede serendipity [9], divergent behaviour that leads to information encounters [7], and feelings of frustration at work that prompt follow-up on an anomalous observation [9].

The precipitating conditions that converge to facilitate or hinder serendipity are just beginning to be understood, but an understanding of these factors is critical to the development of digital environments that support serendipity. Given this, how can we better understand how to support serendipity in a digital environment? I hypothesize that serendipity in information search systems may be facilitated through the design of a system sensitive to external and internal context. Some precipitating conditions are hypothesized to exert more influence on the process of serendipity than others: knowledge, motivation, behaviour, task, environment, and other people. Of these, knowledge, motivation, and behaviour may be observed and measured while task, environment, and other people may be controlled and manipulated.

3 Methods and Sketch of Proposed Solution

My research is being conducted in two main stages and with multiple methods:

Phase 1: My objective is to understand the process of serendipity in knowledge work and specifically a better understanding of the precipitating conditions. To date, I have conducted semi-structured interviews with 12 knowledge workers from a variety of fields such as journalism, education, and science. Transcribed interviews were coded using NVivo. Analysis suggests that while precipitating conditions drive serendipity forward, the context is not always positive. There are multiple factors that affect how we move through our days and these factors push and pull us in different directions. For example, similar to Barber and Fox's [9] observation, a frustrated individual may be more aware of anomalous information, more motivated to make unexpected associations and follow-up on them in order to satisfy a current problem or task.

Phase 2: The objective of the second phase is to understand which precipitating conditions enable people to break from normal thought processes in unexpected ways in a digital environment. While there have been some attempts to support serendipity through useful/surprising embedded links [10] and an ambient intelligence system that draws attention to interesting/surprising content [11], there have been few studies exploring which factors of the external context are more likely to encourage serendipity. In this stage I plan to observe participants using different and contrasting information search systems. In this experiment, the interfaces will be tested using a between and within subjects study design. Motivation and knowledge will be measured while user behaviour will be observed. Participants will be given directed and undirected tasks to test their interactions with the systems and compared. Post session interviews will provide a source for understanding information interactions. One possible direction with regard to interface design is the inclusion of a Twitter widget with tweets engineered to relate to the user's search context. Another possible direction is the development of a scale to measure the user's perception of the potential of a system to facilitate serendipity.

4 Expected Contributions

The results of my Ph.D. research will provide guidelines for the development of digital environments to support serendipity. Namely, this research will indicate ways in which some precipitating conditions such as task and other people interact in a digital information search system to facilitate or hinder serendipity. The main goal and potential outcome of my research is to foster creativity in information interaction such that novel connections are made and useful and interesting paths pursued.

Acknowledgments. Research supported by Social Sciences and Humanities Research Council of Canada (SSHRC), GRAND NCE, CRC, and CFI. Special thanks to my supervisor, Elaine Toms, for her support and guidance.

References

1. Merton, R.K.: The Bearing of Empirical Research upon the Development of Social Theory. Am. Sociol. Rev. 13(5), 505–515 (1948)
2. Cunha, M.P.E.: Serendipity: Why Some Organizations are Luckier than Others. FEUNL Working Paper No. 472. SSRN (2005), http://ssrn.com/abstract=882782
3. Kjölberg, J.: Serendipity in Technology and Education. In: Rogala, W., Selander, S. (eds.) Technology as a Challenge for School Curricula, vol. 11, pp. 1–6. Stockholm Institute of Education Press (HLS förlag), Stockholm, Sweden (2003)
4. Foster, A., Ford, N.: Serendipity and Information Seeking: An Empirical Study. JDoc 59(3), 321–340 (2003)
5. Roberts, R.M.: Serendipity: Accidental Discoveries in Science. Wiley, New York (1989)
6. Heinström, J.: Psychological Factors Behind Incidental Information Acquisition. Libr. Inform. Sci. Res. 28(4), 579–594 (2006)
7. Björneborn, L.: Serendipity Dimensions and Users' Information Behaviour in the Physical Library Interface. Inform. Res. 13(4) (2008)
8. McCay-Peet, L., Toms, E.G.: The Process of Serendipity in Knowledge Work. In: Proc. of the 3rd Symp. on Information Interaction in Context (IIiX), pp. 377–382. ACM, New York (2010)
9. Barber, B., Fox, R.C.: The Case of the Floppy-Eared Rabbits: An Instance of Serendipity Gained and Serendipity Lost. Am. J. Sociol. 64(2), 128–136 (1958)
10. Erdelez, S.: Investigation of Information Encountering in the Controlled Research Environment. IP & M 40(6), 1013–1025 (2004)
11. Beale, R.: Supporting Serendipity: Using Ambient Intelligence to Augment User Exploration for Data Mining and Web Browsing. Int. J. Hum.-Comp. St. 65(5), 421–433 (2007)

Human-Computer Interaction and Human Mental Workload: Assessing Cognitive Engagement in the World Wide Web

Luca Longo

Department of Computer Science and Statistics,
Trinity College Dublin
llongo@cs.tcd.ie

Abstract. Assessing the cognitive engagement of a user while seeking and consuming information over the World Wide Web is a key challenge for studying the quality of interactions. Indicators of cognitive engagement are useful for enhancing usability of interfaces, designing adaptable systems but also for analysing user behaviour and performance. For this purpose, we aim to adopt the multifaceted concept of Human Mental Workload, mainly applied in psychology and cognitive sciences, to study individual performance and user engagement in the context of Web. We aim to design a framework in which mental workload can be measured, analysed and explained. This will lead to analysis of individual and mass behaviour, guidelines and recommendation for interaction design, usability of front-end web applications and proposal of adaptive systems.

Keywords: Human-Computer Interaction, Human Mental Workload, User Engagement, Artificial Intelligence, Web-mining, Human Factors.

1 Introduction

The proliferation of web applications and the explosion of the Internet as an accessible data resource is rapidly changing the way users access, publish, seek and consume information, and knowledge working in general. Interacting with Internet technologies and surfing the WWW is a complex cognitive activity that involves visual, auditory and tactile human modalities rather than physical channels: in these activities human cognition is engaged at multiple stages. Analysing user behaviour while seeking and consuming information over the Web and designing user-centered web applications are becoming key topics in the last few years in the field of HCI. As a consequence, indicators of cognitive engagement are increasingly needed for studying the quality of the user experience while interacting with web resources.

Each action a person performs over the Web can be routinely monitored and stored in web-interaction logs. Existing methodologies for knowledge discovery from such sources are mainly focused on statistical and quantitative analysis. They mainly deal with data such as requested URLs, time properties, the provenance, the type of operating system and browser of a web-user. However, a more fine-grained set of web

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 402–405, 2011.

activities such as clicking, scrolling, keyboard usage, mouse movements, that better describe the user's physical interaction with web interfaces, can now be gathered using new Internet technologies.

Although logs turn out to be detailed sources of information, prior research, both academic and industrial, did not prove useful in interpreting logs from a cognitive perspective that links objective web-actions to cognitive theories of task execution. Reading and interpreting interaction logs, from this perspective, is a non-trivial problem: it is extremely difficult to identify users' original intention, goal, degree of attention and interest from these footprints.

Our proposal is to develop an executable framework, inspired by the concept of Human Mental Workload, large-scale applicable in the context of Web, to measure, assess, study and explain users' attention and cognitive engagement [1]. Models of human mental workload, in other fields, have been proven useful in the design of adaptive interfaces and in the study of individual performance. However, the explicit and invasive way of conducting experiments towards indexes of human mental workload, generally laboratory-driven, have not had a large-scale success.

2 Research Question

To what extent can the Human-Computer(Web) interaction be automatically captured and interpreted towards automatic assessments of Human Mental Workload?

What are the factors that contribute to define human mental workload in the World Wide Web.

3 Approach

Our aim is to define a framework in which the concept of human mental workload can be measured, analysed and explained. The adopted approach is holistic and convergent, and it follows a general progression:

- analytical/empirical exploration of human mental workload characteristics;
- resolving and correlating these characteristics to reliable objective measures;
- definition of an extensible context-aware computational model for the Web;
- validation of the designed prototype with reliable subjective measures.

The research agenda includes the following stages:

1. implementation of a technique that allows non-invasive monitoring of web-users' activity without influencing their experience while consuming information over the World Wide Web;
2. implementation of a technology, large-scale applicable, capable of gathering a more fine-grained corpus of interaction fingerprints such as clicking, scrolling, mouse movements and keyboard usage, closer to the physical human-computer interaction;
3. identification of interaction patterns and development of human mental workload characteristics for the Web;

4. designing an extensible computational model for aggregating human mental workload characteristics;
5. validation and justification of generated indexes of human mental workload by comparison with existing psychology and cognitive measures [1, 2];
6. application of generated indexes for designing engaging technologies and web applications.

The first 2 stages are aimed at developing a transparent non-invasive technology for gathering detailed human-computer interaction patterns while surfing the Web. The third aims to identify and objectively assess the factors that influence human mental workload in the context of Web. The fourth includes the design of a computational prototype, scenarios-aware, able to aggregate the factors individuated in 3 for a certain web-scenario. The fifth stage is aimed at validating the computational model via comparisons with subjective existing measures, successfully applied in psychology or cognitive science. The last gives practical means to deploying metrics, definitions and techniques for designing adaptable and engaging web applications. The research stages are not independent, rather they are related aspects that inform and influence the entire approach towards the measurement and definition of human mental workload The ongoing work is at stage 3 and 4. The monitoring technology has been deployed and the human mental workload characteristics along with the parallel definition of the computational model, are investigated.

4 Sketch of the Solution

Our framework is composed by 5 functional-based macro-levels. The *human-computer interaction level* underlines the process in which users interact with web-interfaces seeking and consuming information. Their interaction is monitored at the *technology level*: a designed technology aims to mine detailed signs of interaction such as mouse movements, keyboard usage and web-browser events over time. Here a raw interaction log is produced and subsequently consumed at the *pre-processing level*. Row data is pre-processed using machine learning techniques towards the identification of web-interaction sessions and tasks within each session. Each task is then recognised to be of a certain typology, thanks to an abstract high-level task-taxonomy. This carries information about task demands in terms of cognitive resources and difficulty. At the *computational heuristic level* characteristics of human mental workload are individuated, computed and aggregated towards a unique index for a given task. Example of characteristics that may influence the mental workload are task interference, cognitive resources overlapping between multiple tasks, task concurrency, time sharing as suggested by the multiple resource theory of attention proposed by Wickens [3, 4]. Individual skills and traits are important characteristics that undoubtedly moderate levels of workload as well as time and richness/control of online activity. The human mental workload characteristics relate to specific scenarios and how they can be calculated by adopting state-of-the-art metrics in psychology, IR or novel metrics. The last *application level* is represented by the practical use of indexes of human mental workload towards the design of user-centered technologies and user engagement aware web-applications.

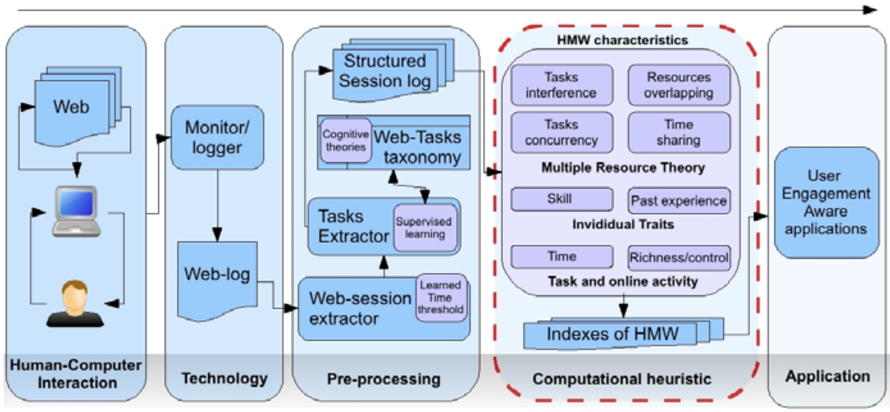


Fig. 1. Anatomy of a framework based on Human Mental Workload assessments

5 Expected Contribution of the Research

The research will contribute to the field of Human-Computer Interaction within several areas, proposing a novel large-scale executable paradigm capable of human mental workload automatic assessment in the context of Web. The expected outcomes of the line of research are to produce an applicable framework able to lead to generalisable design recommendations and strategies along with examples and guidelines employable by other researchers or designers. Our contribution will provide reference and case studies for the application of a theory useful for promoting user-engagement aware applications and interactive design contributing to the appreciation and support of design practice in HCI.

References

1. Kantowitz, B.H.: Development of Nasa-TLX (Task Load Index): Results of Empirical and Theoretical Research. Human Mental Workload 52, 139–183 (1988)
2. O'Brien, H.L., Toms, E.G.: What is user engagement? A conceptual framework for defining user engagement with technology. Journal of the American Society for Information Science & Technology 59(6), 938–955 (2008)
3. Wickens, D.C., McCarley, J.S.: Applied Attention Theory. CRC, Boca Raton (2008)
4. Wickens, C.D.: Multiple Resource and Performance Prediction. Theoretical Issues in Ergonomics Science 3(2), 159–177

Human-Computer Interaction for Security Research: The Case of EU E-Banking Systems

Caroline Moeckel*

Faculty of Business, Department of Informatics,
London South Bank University,
103 Borough Road, London, SE1 0AA, UK
moeckelc@lsbu.ac.uk

Brief Description. This short paper highlights the interaction between security and usability in e-banking security and presents objectives and difficulties for studies in this field.

Research Area: Human-Computer Interaction for Security (E-Banking).

1 The Role of Security and Usability in Current EU E-Banking

E-Banking has undoubtedly become a key element in every modern bank's service portfolio during the last two decades. With its initial focus primarily on transactions, e-banking has now progressed towards an advanced online sales channel for financial products and most of today's e-banking platforms will be able to reproduce almost all services offered in bank branches [1]. But not only the strategic and commercial sides of e-banking have developed very far, also the e-banking security countermeasures employed have passed through several development stages, e.g. from simple password systems to sophisticated two-factor authentication approaches [4].

While the described innovation should generally be viewed as a positive one, current e-banking systems still show large failings in terms of their security and usability. These weaknesses are for example evident in a persistently high number of fraud cases for online banking, accounting for £46.7 million in 2010 (2009: £59.7, 2008: £52.5 [2]) solely in the UK [3]. The risk posed by banking fraud may also lead to a growing negative public perception or fear shown by customers. Alongside these security issues, some severe usability drawbacks can be observed in modern e-banking systems. While some of these may be related to the actual user interface design, a range of problems arise at system stages such as registration or business continuity. System lock outs or erroneously blocked transactions, like Barclaycard's fraud detection system flagging payments for London 2012 Olympic tickets as suspicious and declining them, are incomprehensible and frustrating to users. There are a range of similar cases, such as mentioned in [5] where card reader usability impedes security or in [6] where users prefer less secure solutions due to their apparently higher usability.

* Supervised by Xristine Faulkner. Email: xristine@lsbu.ac.uk

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 406–409, 2011.

Security problems in this context may lead to substantial monetary losses or damage to reputation. However, usability shortcomings may also result in higher operational costs, for example when additional customer support is required through call centres or customers stop using the e-banking functions offered and revert to branch-based services. Furthermore, customers may prefer other financial institutions if their e-banking system seems to be more usable, e.g. if the bank forces the user to purchase a certain type of card reader. Due to the relative importance of e-banking for financial online product sales [1], usability may also directly affect revenue.

Considering the impact of security and usability as well as the examples and fraud statistics mentioned, the difficulty of combining the roles of security and usability in the context of e-banking becomes obvious. Although modern e-banking systems certainly provide sufficient protection levels and fulfil most functions expected from an internet-based bank branch, there is still a significant potential for improvement in the field of usable security. While the best possible protection against potential attacks and mitigation of threats is desirable from a security point of view, the proposed solution still has to be as usable as possible, which may ultimately require accepting a trade-off in security [7]. A range of security and usability problems in e-banking systems can be attributed to a conceptual mismatch of threats and countermeasures within the individual system - either at the cost of security when threats are not mitigated adequately by the employed countermeasures or at the cost of usability due to overly strict, hindering security measures not in line with the level of risk encountered. In addition, usability is often viewed as a separate entity and not considered at every stage of development, which may ultimately lead to decreased security or usability levels. Factoring in risk is crucial at this point as the risk level willingly accepted by a bank will influence the selection and design of their security countermeasures. Accepting certain calculated risks in this context may also be of benefit for usability, for example when allowing users to view their account balance after login in with a username-password combination and only enforcing complex security measures for higher-risk services such as transactions. Viewing security and usability purely as antagonists would be too one-dimensional [8], it is their interaction that needs to be understood and further defined for improving e-banking security.

There is a high degree of freedom in selecting countermeasures and designing systems in the context of e-banking and only a limited amount of regulatory guidelines on the issue. This is also underpinned by a large variation of different e-banking security systems currently observed across the European Union (EU) [1]. Although financial institutions are protecting themselves and their customers against a common global threat landscape, the systems currently in existence in the EU vary in terms of countermeasures employed and their respective usability.

Banks as well as academic researchers would therefore greatly benefit from more insights into the relationship between usability and security in the context of e-banking, particularly if these results could be translated into guidance applicable to real-world systems. Research in the area of usable security is generally limited due to the relative youth of the discipline [8]. In addition, there have been no academic or commercial efforts to create a comprehensive overview of the current e-banking security landscape in the European Union.

2 Benefits and Objectives of Research in the Field

To enhance usability in e-banking security, overcome current usability or security flaws and to ultimately improve e-banking systems, the understanding of the relationship between security and usability in this context needs to be further deepened, formalised and related to a real-world context. A lack of dedicated research methods for the field of usable security as well as the highly diverse sample of e-banking security solutions with different levels of security and usability in the current EU banking landscape indicate the potentially high value of a study in this area.

For research in this field, the following three main objectives can be identified:

1. creating a comprehensive overview and categorisation of e-banking security systems currently in real-world use throughout the EU,
2. understanding the relationship between security and usability in the context of e-banking as well as the potential influence of other factors such as cost and
3. extending the knowledge in the field of research methods for usable security as well as threat modelling with an applied example (e-banking).

3 Research Methods for Usable Security in E-Banking

For the case of security research, it needs to be understood that most organisations will not allow external researchers to access all details related to their security policies, specific solutions or products. This applies particularly to the very secretive banking sector, where publicised security flaws can translate into substantial financial losses, decreased trust or reputational problems. The expectation to work with real-world data on security within the organisation, e.g. number of security breaches encountered, investment on e-banking security or plans for future implementations, is therefore not entirely realistic. Ways to overcome this difficulty include using system analysis, e.g. of e-banking applications, customer or employee interviews and questionnaires or using data collected by professional or governmental organisations.

Secondly, research methods in the field of usable security have not fully matured yet and while various key publications [7] [8] in the area have hinted towards certain approaches such as threat modelling [9][10] or user testing, these need to be extended further and tested in various applied contexts. It is argued here that current usability evaluation methods do not fully account for the special nature of secure applications and software. This applies in particular for the example of e-banking, where a highly secure infrastructure has to provide simultaneously a high level of usability to its large number of non-expert users. This view is supported by key materials in this area, which criticise the adoption of usability methods by researchers in human-computer interaction for security (HCISec). Kainda et al. [7] notes that "the extent to which these apply to the field of HCISec is arguable given the fine balance between improving the ease of use of a secure system and potentially weakening its security". Additionally, Cranor et al. [8] highlight the importance of "contextual information [...] in assessing the cost of the countermeasure to the system as a whole—this includes financial, organizational, and user costs". This is by no means to say that usability evaluation methods should not be employed in a security context, since they may be beneficial through improving users' effectiveness, efficiency or satisfaction, but hints towards an integration of threats and vulnerabilities and surrounding factors.

4 Original Contribution and Future Work

A number of innovation factors valuable to banks will be included in this study. A framework for evaluating existent and future implementations of e-banking security will be devised for practical employment. Furthermore, the bank's position with all its related challenges will be taken into account rather than assessing the situation in a theoretical, unrealistic environment. In contrast to earlier studies, the interplay of security and usability without a separation of factors will be analysed, enabling the bank to understand the effects of changes in security or usability. In addition to introducing a new focus of user-centered design to e-banking security, this study will also give a detailed overview of the status of EU e-banking security.

Future work in this area could include exploring the mismatch between the customer's and the bank's perspective on e-banking security, further examining the data collected on e-banking security in Europe, learning more about the internal reasons that banks prefer certain countermeasures over others as well developing usability evaluation models specific to e-banking security.

References

1. Moeckel, C.: EU B2C E-Commerce in the Banking Sector. Diplomica, Hamburg (2008)
2. Financial Fraud Action UK: Fraud the Facts (2010), http://www.financialfraudaction.org.uk
3. UK Cards Association: Fraud losses drop on UK card, cheques and online banking (March 2011), http://www.theukcardsassociation.org.uk/media_centre/press_releases_new/page/1323
4. Borchert, B.: Online-Banking Verfahren, http://www-ti.informatik.uni-tuebingen.de/~borchert/Troja/Online-Banking.shtml
5. Drimer, S., Murdoch, S.J., Anderson, R.J.: Optimised to Fail: Card Readers for Online Banking. In: Dingledine, R., Golle, P. (eds.) FC 2009. LNCS, vol. 5628, pp. 184–200. Springer, Heidelberg (2009)
6. Weir, C., Douglas, G., Richardson, T., Jack, M.: Usable security: User preferences for authentication methods in eBanking and the effects of experience. Interacting with Computers 22(3), 153–164 (2010)
7. Kainda, R., Flechais, I., Roscoe, A.W.: Security and usability: analysis and evaluation. In: 5th International Conference on Availability, Reliability and Security, pp. 275–282. IEEE Press, New York (2010)
8. Cranor, L.F., Garfinkel, S.: Security and Usability - Designing Secure Systems That People Can Use. O'Reilly, Sebastopol (2005)
9. Moeckel, C., Abdallah, A.E.: Threat Modeling Approaches and Tools for Securing Architectural Designs of an E-Banking Application. In: 6th International Conference in Information Assurance and Security, p. 149. IEEE Press, New York (2010)
10. Moeckel, C., Abdallah, A.E.: Understanding the Value and Potential of Threat Modeling for Application Security Design: An E-banking Case Study. Journal of Information Assurance and Security 6(4) (to appear, 2011)

Information Architecture Automatization for the Semantic Web*

Josep Maria Brunetti and Roberto García**

Universitat de Lleida,
Jaume II, 69,
25001, Lleida, Spain
{josepmbrunetti,rgarcia}@diei.udl.cat

Abstract. Our proposal is to develop generic Information Architecture components to facilitate publishing and browsing semantic data in the web, improving its usability and accessibility.

Keywords: Semantic Web, Linked Data, Information Architecture, Usability, Accessibility, Navigation, Metadata.

1 Introduction

Despite the Semantic Web was proposed ten years ago [1], it hasn't been until recently when it has started to become popular. In the last years, the amount of semantic data available in the Web has increased in a spectacular way, especially thanks to initiatives like Linked Open Data (LOD). The objective of this initiative is to motivate the publication of Open Data in formats that are more easily integrable, queryable and that facilitate its reuse. The cloud of interrelated and open datasets included in the LOD cloud has rapidly evolved, from the 2 billion statements and 30 datasets one year after its creation in February 2007, to more than 25 billion statements and 200 datasets in September 2010[1].

The potential of this huge amount of data is enormous but in most cases it is very difficult for users to explore and use these data. The problem is that most of this data is available as raw data dumps or SPARQL [2] semantic query services, being it complicated to realise what the data is about, which are the main kinds of things, how are they interrelated, etc. Moreover, it requires some experience in Semantic Web tools and, in any case, the results are not very usable. Therefore, the challenge is to structure all this information making it more usable and accessible.

Information Architecture (IA) is a discipline that focuses its efforts in this problem, especially in complex systems and situations with great amounts of information. IA describes the structure of a website, how its information is grouped, navigation

* The work described in this paper has been partially supported by Spanish Ministry of Science and Innovation through the OMediaDis research project (TIN2008-06228).
** This dissertation work is supervised by Roberto García.
[1] http://linkeddata.org/

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 410–413, 2011.

methods and terminology used. A good IA can improve the quality of a website and users can find more easily the information they are looking for.

The proposal of this project is to draw from the experience accumulated in the Information Architecture domain [3] and adapt existing IA components to these situations with large amounts of heterogeneous semantic data. The rest of this paper is structured as follows. In Section 2 we present the state of the art and other approaches to this problem. Section 3 is about our proposed approach and in Section 4 we explain the methodology we will follow during this PhD and evaluation methods. Finally, in Section 5 we present our conclusions and future work.

2 State of Art

The best approach to make a semantic dataset more usable to a wider range of users is to use a Web data publishing tool or a Semantic Web browser, e.g. Disco [4]. These kinds of tools provide an HTML rendering for each resource in the dataset, listing all their properties and values. HTML pages are interlinked based on the connections between resources in the underlying dataset. However, this feature is only useful if the user has some a priori knowledge of the data structure. There is no way to get an overview of the kinds of resources in the dataset, e.g. person, address, etc.

Explorator [5] is another tool that makes it possible to browse a dataset combining search, facets and operations on sets of resources. However, it is also difficult for users to get a broader view on the dataset. Other tools also provide faceted views but not as a generic browser, i.e. just for a specific dataset like the case of the DBPedia Faceted Browser [6] or \facet [7].

Consequently, existing tools make very difficult for users to explore a dataset, realize what kind of resources there are, what properties they have and how they are related.

3 Proposed Approach

As the volume of information in the Semantic Web increases, interacting with information becomes a more difficult task. To interact with these amounts of data, users use different ways or strategies depending on their goals.

Starting from the fundamental set of tasks for data analysis proposed by Schneiderman [8], we have explored the most appropriate Interaction Patterns to perform these tasks in the context of the Semantic Web:

1. **Overview:** The goal of this task is to get a full view of the entire collection. We propose to apply the Global Navigation interaction pattern. In the context of IA it corresponds to **navigation menus**.
2. **Filter:** To select items of interest and filter out uninteresting items through exploratory search. The proposal is the Faceted Navigation pattern, which in the context of IA corresponds to **facets**.
3. **History:** To keep a history of actions and support undo and replay actions. We propose to apply the Breadcrumb Navigation pattern. In the context of IA it corresponds to **breadcrumbs.**

We have chosen these IA components because they are simple and very common in websites, so users are comfortable with them. Besides them, we are exploring other IA components such as **sitemaps** and **search systems**.

The drawback of all these IA systems is that they are quite expensive to develop and maintain, even more with the large amount of heterogeneous semantic data we are dealing with. Fortunately, when these IA systems are built on top of structured data and formal languages, it is possible to automate most of this development and maintenance work.

Concretely, the Semantic Web, with the use of ontologies and semantic languages like RDF, allows the automatization of these components. The Semantic Web provides methods and tools to model the Information Architecture for a domain with much more detail and in a formal way. It is possible to define concepts, properties and relations between them. This allows the use of automated tools to process these formal descriptions, create and maintain the Information Architecture.

The main goal of this PhD is to define a methodology and develop the necessary components to assist during the definition and maintenance of the Information Architecture of websites from a formal description of the site's domain. It is also intended, in parallel, to improve the usability and accessibility of the generated Information Architecture. By using these components, the final users will be able to easily explore semantic data, which at this moment is not usable and accessible for them.

4 Methodology and Evaluation

The methodology to follow in this project is based on the MPIu+a [9] development process, which integrates the disciplines of Software Engineering, Human-Computer Interaction and Accessibility. Besides the traditional Software Engineering lifecycle, MPIu+a has two other main blocks, prototyping and evaluation, which are also two main parts of this project.

We plan to develop prototypes for each of the identified Information Architecture components, test and evaluate them with users. The evaluation will be conducted with common evaluation methods such as user tests, card sorting or heuristic evaluation. The objective of the evaluation is to test and compare the generated Information Architecture with a traditional one.

Although this project has just barely started, a prototype[2] has already been tested with end-users in order to evaluate its functionality and usability. The goal of the tests conduced so far was to do a preliminary evaluation of the menu and facet components, to check if they are understood and they improve user performance when looking for a specific piece of information. The results so far show that LinkedData can improve the user experience, providing more ways to interact with data [10].

5 Conclusions and Future Work

In this paper we have presented our vision of this doctoral thesis in the context of the Semantic Web and Information Architecture. We have started developing generic IA

---

[2] `http://rhizomik.net/linkedmdb/`

components that can be reused in websites with totally different domains as long as they are based on Semantic Web technologies.

The identified IA components are able to use the semantics captured by ontologies and semantic data, providing users different ways to access and interact with the data. Those components, which are automatically generated and maintained, facilitate publishing and browsing a dataset without requiring a priori knowledge of it or experience in Semantic Web tools.

The short-term future work focuses now on implementing the improvements detected during the user tests, basically those that affect the navigability between resources in the dataset. The main problem detected is that the user interaction is currently too constrained by how the underlying data is structured. The objective is to improve IA flexibility and obtain implicit properties and relations, for instance reverse properties, in order to provide users with alternative paths.

Then, we also plan to develop prototypes for the other identified IA components, test and evaluate them with users.

References

1. Berners-lee, T., Hendler, J., Lassila, O.: The semantic web. Scientific American 284, 34–43 (2001)
2. Prud'hommeaux, E., Seaborne, A.: SPARQL Query Language for RDF. Recommendation, World Wide Web Consortium (2008), `http://www.w3.org/TR/rdf-sparql-query`
3. Morville, P., Rosenfeld, L.: Information Architecture for the World Wide Web. O'Reilly Media, Cambridge (2006)
4. Bojars, U., Passant, A., Giasson, F., Breslin, J.G.: An Architecture to Discover and Query Decentralized RDF Data. In: Proceedings of Workshop on Scripting for the Semantic Web, SFSW 2007. CEUR Workshop Proceedings, vol. 248 (2007)
5. Araujo, S., Schwabe, D., Barbosa, S.: Experimenting with Explorator: a Direct Manipulation Generic RDF Browser and Querying Tool. In: Visual Interfaces to the Social and the Semantic Web (VISSW 2009), Sanibel Island, Florida (2009)
6. Hahn, R., Bizer, B., Sahnwaldt, C., Herta, C., Robinson, S., Bürgle, M., Düwiger, H., Scheel, U.: Faceted Wikipedia Search. In: Abramowicz, W., Tolksdorf, R. (eds.) BIS 2010. Lecture Notes in Business Information Processing, vol. 47, pp. 1–11. Springer, Heidelberg (2010)
7. Hildebrand, M., Ossenbruggen, J., Hardman, L.: /facet: A Browser for Heterogeneous Semantic Web Repositories. In: Cruz, I., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M., Aroyo, L.M. (eds.) ISWC 2006. LNCS, vol. 4273, pp. 272–285. Springer, Heidelberg (2006)
8. Shneiderman, B.: The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In: Proceedings of the 1996 IEEE Symposium on Visual Languages (VL 1996), pp. 336–343 (1996)
9. Granollers, T.: User centred design process model, integration of usability engineering and software engineering. In: Proc. INTERACT 2003, Zurich, Switzerland, pp. 673–675 (2003)
10. Gil, R., Brunetti, J.M., López-Muzás, A., Gimeno, J.M., García, R.: Improved Linked Data Interacting with the Web of Data through Information Architecture Components. In: INTERACT 2011 (submitted, 2011)

Microinteractions to Augment Manual Tasks*

Katrin Wolf

Deutsche Telekom Laboratories, TU Berlin, Ernst-Reuter-Platz 7
10587 Berlin, Germany
katrin.wolf@acm.org

Abstract. This paper summarizes the first nine months of progress on my Ph.D. project. The research focus of the project is on investigating microinteractions. a sub-topic of HCI and gesture research. The project will develop a framework for applications that use microgestures to support everyday tasks through invisible and context-aware appearing interface layers underneath object-grasping hands. In an expert study that has been accepted as a full paper at INTERACT 2011, I explore the motor limitations and opportunities of microgestures while grasping objects and valued manual dual-task scenarios by walking through three tasks that involve grasping objects. The outcome of the study is a generic microgesture set for different grasp types and a collection of parameters that have a relevant effect on the choice of the grasping tasks. A further user study in progress is investigating the effect of grasped objects, such as handheld devices, on the feasibility of performing microgestures. Users are asked to perform finger-tip and drags on the front and/or back of a handheld device. The device is two-sided and touch-sensitive, it is made by stacking 2 pads together in a sandwich-like prototype. This allows tracking users' finger gestures through a camera as well as through front and touch screens. The outcome of the two mentioned studies will describe a design space for out-of-a-grasp microgestures. At the INTERACT doctoral consortium I aim to present this design space and discuss how this can serve as a basis for developing a framework of out-of-grasp microinteractions that are subtasks of grasping tasks. The microinteractions will be developed to support the grasp tasks with regard to their perceived ergonomic and hedonic qualities.

Keywords: Microinteraction, gestures, dual-task, multitask, interaction style.

1 Research Focus

Human actions in the real world are usually a set of tasks solved in parallel and people have the ability to multi-task with respect to motor and cognitive resource handling in many everyday situations. Normally, human-computer interactions are designed as separate tasks and synchronous tasks are understood as multitasking

---

* This dissertation work is supervised by Prof. Sebastian Möller (Quality and Usability Laboratories, TU Berlin, sebastian.moeller@telekom.de) and co-supervised by Michael Rohs (LMU Munich), Anja Naumann (TU Berlin), and Joachim Sauter (UdK Berlin / ART+COM).

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 414–417, 2011.

situations with two competing tasks when considering their motor and cognitive resource requirements. Microinteractions are a promising technique that can be done in parallel to manual actions with less competitive motor resource effort.

Ashbrook defines microinteractions as short-time interruptions of primary tasks for controlling mobile devices [1]. These interaction techniques have the beneficial potential of allowing mobile application control in parallel with ongoing everyday tasks. This could significantly enrich the quality of tasks that we could perform on-the-go and fundamentally alter the way we view ubiquitous computing [6]. In this context a microinteraction can be designed as a support or subtask of an everyday task or contextual independent, like mobile communications.

I aim to investigate the potential of microinteractions that are a subtask of a primary interaction. In this field several research projects have recently been conducted to develop tracking interfaces [1, 2, 5, 6] and a few projects have been conducted that are driven by the human side of interaction [3, 4, 11]. I see a research gap on the human-centric side of microinteractions, my motivation is to investigate the design space of these interactions; mainly their affordances and constraints regarding usability aspects, such as ergonomic and hedonic qualities. This work will guide the next step of my research, developing a framework for sub-tasking microinteractions that support grasp tasks. Subtasks in this context could include automotive control interactions such as stopping a turn signal or opening the window while driving. Realizing these subtasks through microgestures allows for performing them in parallel to steering a car while keeping the steering wheel fully grasped.

I believe that synchronous microinteractions allow for the possibility of supporting every-day tasking in an ergonomic and hedonic manner. Therefore my hypothesis is: Microinteractions have the potential to increase the perceived ergonomic and hedonic qualities of contextually related tasks that are performed in parallel.

2 Understanding Microinteractions

To help understand microinteractions, I did an expert evaluation [10]. I asked sports therapists and physiotherapists to use props (as shown in figure 1) while performing given microgestures, but also they could use any new gestures that they created spontaneously. The gestures should be easily performable without interrupting the primary task, without needing high cognitive effort, and without taking any risk of being mixed up with natural movements.

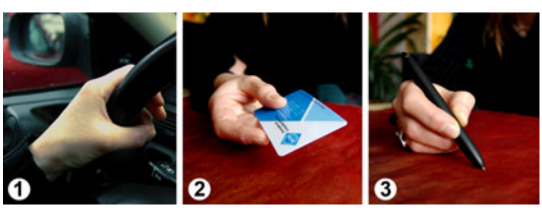


Fig. 1. The participants are testing the feasibility of finger gestures while (1) holding a steering wheel (2) targeting a cash card, and (3) drawing with a pen

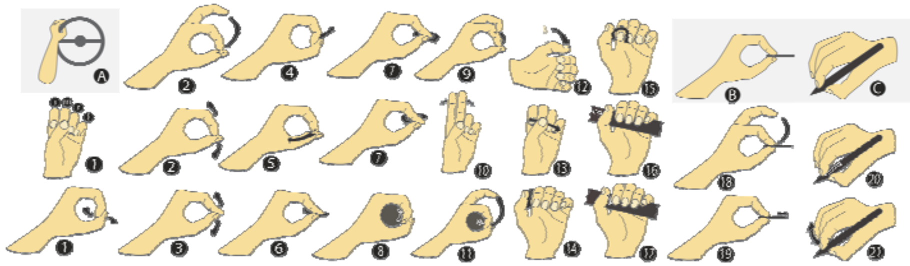


Fig. 2. The expert-defined and evaluated gesture set. The experts found 17 gesture types for the driving scenario (A). The card targeting scenario (B) and the stylus scenario (C) just contain 2 gesture types each. Most gesture types have several sub-types by performing them with different fingers (index, middle, ring, and little finger). Moreover the same gesture results in a different sub-type (e.g. touch, tab, or press), if it is performed with different acceleration or duration.

From the expert interview results I developed a taxonomy (see figure 2) for classifying these gestures according to their use cases and assessed their ergonomic and cognitive attributes, focusing on their primary task compatibility. I defined 21 gestures, which allow for microinteractions within manual dual task scenarios. In expert interviews I evaluated the gesture's level of required motor or cognitive resources under the constraint of stable primary task performance.

This study was supervised by three researchers (one psychologist and two computer scientists) of the research group I am associated with at the Quality and Usability Laboratories of TU Berlin and that are named as co-authors.

3 Research Plan and Status Quo

My Ph.D project is planned to consist of three parts and I am currently working on the first one. This part investigates the fundamental understanding of microinteractions. Within the second part, I will – based on the expertise of part 1 – develop novel microinteraction techniques and evaluate them regarding their usability, such as ergonomic and hedonic qualities. The third part will be dedicated to develop a framework for microinteractions that will serve as seamless sub-tasks to support in parallel continued grasp-based tasks.

My first experiment [10] serves as an initial analysis basis for understanding microinteractions and will be continued to generate a fundamental description of the design space for out-of-grasp microgestures. In this initial study, attributes that have a relevant effect on the gesture performance are defined. These are the grasp type, the grasped objects, ergonomic and bio-mechanic aspects, as well as the cognitive effort that is required for the pure gesture performance. The benefit of microinteractions on the usability of the grasp tasks was valued by experts as useful augmentation for many grasp-based tasks but they also brought up the fact that the degree of benefit might be influenced by pragmatic aspects, such as the task length and by ergonomic parameters of the grasped object.

Guided by the conclusion of the expert study, I currently plan a user study to investigate how users grasp an object like a pad and the effect of physical object attributes (use, size, and shape) on the way users grasp it and on their finger gesture feasibility. The expected outcome of this study is a deeper understanding about users' naturally chosen grasp for holding a pad (e.g. iPad) while performing certain device interactions as well as an evaluation of users' ergonomic skills of performing gestures while grasping the device. To measure the limitations of the grasp performance I will choose certain tasks that have to be done under different conditions (with one or two hands / with or without releasing other fingers). A generic set of finger gestures that are performable by a grasping hand are given by the Taxonomy of Microinteractions [10]. To measure the specific gesture limitations while grasping an object, we ask users to perform guided gestures while holding an iPad sandwich.

4 Contribution and Expected Benefit of the Doctoral Consortium

Within the student consortium I will present my previous research [8], [9] and the results of my first study [10] shortly for explaining my current work in greater detail.

In my actual project I measure objective and perceives pragmatic qualities as well as perceived hedonic qualities of microinteractions using SMEQ, NASA TXL, and AttrakDiff. In the student consortium I would appreciate to discuss this approach.

Because I aim to develop a framework for microinteractions, another question to ponder would be how to get valid measurements that are scalable for various grasp-based task types and for grasped objects with different parameters.

References

1. Ashbrook, D.: Enabling Mobile Microinteractions, Doctoral Theses, Georgia Institute of Technology (2010)
2. Li, K., Baudisch, P., Hinckley, K.: Blindsight: eyes-free access to mobile phones. In: Proc. CHI 2008, pp. 1389–1398 (2008)
3. Holz, C., Baudisch, P.: Understanding Touch. In: Proc. CHI 2011, pp. 2501–2510 (2011)
4. Karlson, A., Bederson, B.: ThumbSpace: generalized one-handed input for touchscreen-based mobile devices. In: Baranauskas, C., Abascal, J., Barbosa, S.D.J. (eds.) INTERACT 2007. LNCS, vol. 4662, pp. 324–338. Springer, Heidelberg (2007)
5. Loclair, C., Gustafson, S., Baudisch, P.: PinchWatch: A Wearable Device for One-Handed Microinteractions. In: MobileHCI 2010, Workshop on Ensembles of On-Body Devices, 4 pages (2010)
6. Saponas, T., et al.: Enabling Always-Available Input with Muscle-Computer Interfaces. In: Proc. UIST 2009, pp. 167–176 (2009)
7. Tan, D., Morris, D., Saponas, T.S.: Interfaces on the Go, In XRDS. Crossroads. The ACM Magazine for Students, 30, doi:10.1145/1764848.1764856
8. Wolf, K.: Microinteractions beside ongoing manual tasks. In: Proc. TEI 2011, pp. 447–448 (2011)
9. Wolf, K., Dicke, C., Grasset, R.: Touching the Void: Gestures for Auditory Interfaces. In: Proc. MobileHCI 2010, pp. 305–308 (2010)
10. Wolf, K., Naumann, A., Rohs, M., Müller, J.: A Taxonomy of Microinteractions: Defining Microgestures based on Ergonomic and Scenario-dependent Requirements. In: Proc. INTERACT 2011, 18 pages (2011)
11. Wimmer, R.: Grasp Sensing for Human-Computer Interaction. In: Proc. TEI 211, pp. 221–228

OPEN-HEREDEUX: OPEN HEuristic REsource for Designing and Evaluating User eXperience

Llúcia Masip, Marta Oliva, and Toni Granollers

Department of Computer Science and Industrial Engineering,
University of Lleida (Spain)
{lluciamaar,oliva,tonig}@diei.udl.cat

Abstract. The need of enhancing design and evaluation of user experience exists. We propose a resource which helps to semi automate the design and evaluation of user experience.

Keywords: User experience, evaluation resources, heuristic evaluation.

1 User Experience

The concept of User eXperience (UX) includes more facets than traditional ones related to usability. According to Peter Morville [1], UX includes useful, usable, valuable, desirable, findable, credible and accessible. Moreover, other authors also consider aspects such as communicability, cross-culture, emotions or playability.

Facets help us to improve the quality of use in interactive systems and, consequently, to increase UX. Despite the new tendency in UX, currently there is no specific technique to evaluate UX when users interact with an Interactive System (IS) [2]. Certainly, UX experts can use usability evaluation methods for evaluating UX, because the "oldest" and the most traditional facet used when evaluating aspects related to the quality of use is usability. Nevertheless, a consensus about what is the best methodology to evaluate UX still does not exist [2].

The project which I am involved in with my PhD is contextualized in UX evaluation but focusing our efforts on exploiting the capabilities of one of the methods used most: the heuristic evaluation (HE). HE, as it is known in HCI field, was spread by Nielsen and Molich [3] but HE had been discreetly created by Ravden and Johnson in [4]. It is one of the most challenged and successful methods for evaluating the usability of interactive systems. Nevertheless, HE presents barriers that induce a slower and more expensive evaluation process that could lessen all its advantages. The first deficiency is that the whole evaluation process is based on considered heuristics. A review of this set, to adapt it to specific features of interactive systems is always essential despite the fact that experts have done many efforts to provide a whole set of heuristics and apply them to different interactive systems. The second deficiency is in the evaluation result. Usually, the method provides qualitative results. And more often than not, qualitative results are not enough if these results do not allow us to take decisions in an objective way like it happens with quantitative results. In this context, the following sections explain our ideas to go a step beyond to provide solutions for the mentioned above problems.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 418–421, 2011.

2 OPEN-HEREDEUX

OPEN-HEREDEUX is the short name which we use to refer to the project called: "OPEN HEuristic Resource for Designing and Evaluating User eXperience in interactive systems". It will enable UX experts to design and evaluate UX in a semi automates way. Its four components are detailed below.

2.1 Open Repository of Heuristics

The main goal of this component is to manage information. We should provide a repository with all the information needed to achieve a complete and minimum set of heuristics and to consider all aspects of UX either in its design phase or evaluation process for a specific IS.

Therefore, the repository will have all information such as a whole IS classification [5], essential facets to evaluate UX in any IS and a complete set of heuristics related to all the facets mentioned previously.

Once we have all this information, next step will be data modeling; we will model information through ontologies. Finally, all information will be stored with the most convenient technique to be able to extend, modify or delete this information later. All this information will be used as input for the next component, the heuristics adviser.

2.2 Adviser of Heuristics

The main goal of the heuristics adviser component is to propose the most convenient list of heuristics to be used (for the specific system to be analyzed), such as recommendation principles in a design phase or such as evaluation principles in a UX evaluation based on heuristics.

The adviser considers for its suggestion different factors according to the type of IS and other factors such as economics, etc. About the type of IS the adviser should at least regard the facets considered in this IS and the minimum but complete set of heuristics of each facet according to the user profile that needs these heuristics (designer or evaluator). In reference to economical factors, the heuristics adviser should consider the time that designers or evaluators have to take into account for the set of heuristics and the money invested in the project. These factors should determine the minimum number of heuristics that the adviser should suggest. This component has an important advantage when choosing the best heuristics for a specific IS: turning a manual process and process based on UX expert experience to semi automate process where UX experts will only review the set of heuristics suggested.

Finally, heuristic suggestions can be used either as a list of recommendations to design a specific IS or as an input of a scoring interface.

2.3 Scorer of Heuristics

If we want to use the adviser's set of heuristics to evaluate UX, we can carry out the evaluation using this component. It helps UX evaluators in their scoring of heuristics. It should have a control of user profiles to identify if the user is connected to the system as an evaluation administrator or as an evaluator; the interface provides users with two parts according to each user profile. The first part is the administration part where administrator should review the heuristics suggested by the adviser, and should choose the evaluators and the severity factors. The second part is the evaluation part where evaluators should score every heuristic. In addition, this component stores all the information to make the extraction of qualitative and quantitative results easier.

2.4 Results Analyzer

The last component waits for the extraction of quantitative and qualitative results. In reference to qualitative results we will try to create a list of improvements according to evaluators' observations and heuristics scored incorrectly. They are ones which designers should apply to interactive systems. Apart from qualitative results, we want to obtain quantitative ones because these will try to show the degree of UX that an IS has. If quantitative results are achieved, UX experts will have a standard method or a possible certification to compare evaluations and see which IS provides users with the best experience. Quantitative results will get an objective measure that we cannot achieve with the subjectivity of qualitative results.

3 Research Already Conducted

In reference to the evaluation planning, all these ideas started when we had to carry out an evaluation of physical devices called CIPs (Citizen Information Point) [6]. We realised that we could not use some well-known heuristics such as Nielsen's. These were GUI-focused and, they did not cover our evaluation goals and all the features of CIPs. We started researching as much as possible all general usability heuristics defined since 1986 and we detected 16 repeated categories. They are applicable in simple websites or desktop applications [7].

Apart from this review, we achieved a classification of IS components that permitted us to know what parts of the IS should be assessed and to choose the best heuristics for each part [5]. Then, we validated these 16 categories making a comparison between two sets of heuristics to detect which set can find more usability problems and to detect improvements in their definition [8]. And in addition these heuristics might work in semantic applications, now we are working in it. In reference to the results, we have already made an effort to achieve a mathematical formula in usability HE [8], we do not rule out our option but we know that we have to do more research on it.

4 Current State of Study and Future Research

Currently, we are working on different areas of this project. The first one is modeling dates to be able to achieve the first version of the adviser. Second, we are working on the design of the second component of our resource: the heuristics scorer component. In addition, while all this work is being done, we are also doing research on all the information needed for our open repository although we know this information about heuristics, facets and new interactive systems can increase (for this reason our repository is an open repository).

In the future we will primarily work on quantitative results to achieve a statistical or mathematical formula towards UX standardization.

Acknowledgements. The work has been partially supported by Spanish Ministry of Science and Innovation through the Open Platform for Multichannel Content Distribution Management research project (TIN2008-06228) and it has been partially supported by Universitat de Lleida for pre-doctoral fellowship to Llúcia Masip.

References

1. User Experience design, `http://semanticstudios.com/publications/semantics/000029.php`
2. Vermeeren, A., Lai-Chong Law, E., Roto, V., Obrist, M., Hoonhout, J., Väänänen-Vainio-Mattila, K.: User Experience Evaluation Methods: Current State and Development Needs. In: Proc. of NordiCHI 2010. ACM, New York (2010)
3. Nielsen, J., Molich, R.: Heuristic evaluation of user interfaces. In: CHI 1990: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, NY (1990)
4. Johnson, G.I., Clegg, C.W., Ravden, S.J.: Towards a practical method for user interface evaluation. Applied Ergonomics 20(4), 255–260 (1989)
5. Masip, L., Oliva, M., Granollers, T.: Classification of interactive systems components enables planning heuristic evaluation easier. In: HCI International 2011 (accepted paper, 2011)
6. Oliva, M., Masip, L., Granollers, T.: Evaluación de usabilidad y accesibilidad de un conjunto de dispositivos interactivos denominados Puntos de Información Ciudadana. Scire: representación y organización del conocimiento, num 202212 (2010) ISSN: 1135-3716
7. Masip, L., Oliva, M., Granollers, T.: Hacia la semiautomatización de la evaluación heurística: Primer paso, categorización de heurísticas. In: Interacción 2010, Grupo Editorial Garceta, Valencia, September 7-10, pp. 359–368 (2010)
8. Masip, L., Granollers, T., Oliva, M.: A Heuristic Evaluation Experiment To Validate The New Set Of Usability Heuristics. In: Proceedings of 8th International Conference on Information Technology: New Generations, Las Vegas (2011); 978-0-7695-4367-3/11 © 2011 IEEE DOI 10.1109/ITNG.2011.82
9. González, M., Masip, L., Granollers, T., Oliva, M.: Quantitative analysis in a heuristic evaluation experiment. Advances in Engineering Software 40(12), 1271–1278 (2009)

Sketching Language: User-Centered Design of a Wizard of Oz Prototyping Framework

Stephan Schlögl

School of Computer Science & Statistics
Trinity College, University of Dublin
O'Reilly Institute, Dublin 2, Ireland
schlogls@scss.tcd.ie

Supervisor

Dr. Gavin Doherty
Gavin.Doherty@scss.tcd.ie

Research Area

Wizard of Oz prototyping with modern Language Technology Components;

Research Topic

This research aims at understanding the challenges involved in running Wizard of OZ experiments and searches for an optimal software interface supporting the actions of the wizard.

Research Description

Combining novel speech and language technologies into systems across diverse domains poses significant software engineering challenges. Assuring the usability of the resulting applications is an important, and somewhat under-researched, issue in this area. The use of Language Technology Components (LTC), i.e. Automatic Speech Recognition, Machine Translation and Speech Synthesis, has significantly increased in recent years as their performance has improved. Examples include speech-based interaction in cars to keep a driver's attention on the road and the use of web-based translation tools such as *Google Translate* and *Yahoo! Babel fish*, to understand text that is written in a foreign language.

Due to the fallible nature of the technology involved in such applications, one must ensure that usability and human factors are adequately accounted for. As with applications based on a Graphical User Interface (GUI) software using LTCs needs to be tested early in the design process. While low-fidelity prototypes assessing GUI applications such as sketches and wireframes can be built relatively quick and inexpensively, the development of prototypes evaluating the use of LTCs tends to be both cost and time intensive. An easier and more efficient way of creating those prototypes is desirable. One technique that has been used to test applications involving LTCs is Wizard of Oz (WOZ). In a WOZ experiment a human wizard mimics some of the functionality of a system. Since the technical requirements for

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 422–425, 2011.

such a prototype can be reduced to a minimum, this technique is particularly useful for early stage evaluations and therefore can be seen as a good candidate for addressing the lack of low-fidelity prototyping methods supporting the use of LTCs. An important difference between prototyping GUI based applications and conducting a WOZ experiment is, however, that the latter depends on the actions of the human wizard. Whereas with GUI commands a specific system behavior can be defined in advance by referring to concrete events like mouse-clicks and keyboard-entries, with language technology this clear binding is missing. Here the interaction rather depends on the way a user's input is interpreted. In a WOZ experiment the wizard is responsible for this highly demanding cognitive task [5]. Supporting this role with a usable wizard interface seems therefore crucial.

While several wizard interfaces have been built to date (e.g. [4], [2], [3] and [1]), most of them were designed with specific experiments in mind. The more general issue of understanding the task of the wizard, in order to develop a generic WOZ interface that can be used across different experiments and settings, has remained largely unexplored. The goal of this research is therefore to create a WOZ prototyping framework that aims at a template-based generation of wizard and client interfaces for different application scenarios. Besides providing an environment that offers an easy way of prototyping LTC-based software it is further an important goal of this work to understand and optimally support the task of the wizard.

Research Questions

Based on the stated goals the following research questions arise:

- What is the design space for applications using LTCs and for which of those scenarios can WOZ realistically be used as a low-fidelity prototyping method?
- What are the various aspects of the wizard task and especially what kind of problems does a wizard need to deal with when designing and running an experiment?
- Can the function of the wizard be supported and if yes what features can be implemented into a WOZ prototyping framework that would disburden this task?

Research Hypothesis

The main hypothesis for this research is that by providing an optimal interface for the wizard the complexity of the wizard task can be reduced and that this improvement would increase the wizard's overall performance in terms of consistency and machine-like behavior. It is furthermore assumed that it is possible to structure and classify WOZ experiments, and that this helps to build a software framework, which allows for a somewhat generic creation of wizard and client interfaces.

Research Methodology

In order to fully understand and optimally support the design and conduct of WOZ experiments a User-Centered Design (UCD) methodology is followed. UCD encompasses the following steps that iteratively lead through the planning, design and development of a new product:

- Specify the context of use: This first stage of UCD focuses on identifying the people who will use the future product, the conditions under which it will be used and the reason why it is used. For the proposed research this contextual information will on the one hand be drawn from analyzing the literature, and on the other hand be obtained through talking to wizards and observing their actions.
- Specify requirements: Based on the context of use the second stage of UCD defines requirements that need to be met for the product to be successful. Requirements can be separated into business requirements e.g. the functionality of the product and user requirements e.g. the usability of the product. In the case of WOZ the basic requirements will be derived from existing WOZ tools. Additional aspects will materialize whilst talking to wizards and analyzing their demands.
- Create design solutions: After defining the product requirements, different possible design solutions should be implemented. With the WOZ framework it is planned to employ various prototyping methods that reach from basic sketches on paper to more elaborate functional designs based on web technologies.
- Evaluate designs: As with most product development processes the evaluation phase is crucial for the quality of the product. Testing the different design solutions with actual users ensures appropriateness and applicability of the design and fosters its usability. In order to evaluate the different aspects of the proposed WOZ framework small user studies with different wizards, as well as longer lasting studies that look at the performance of a single wizard over time, are planned.

Research Plan

Up to this point all of the above mentioned steps have undergone at least one iteration. The context of use was explored through testing and comparing existing prototyping products. A literature review highlighted several application scenarios and helped to define concrete requirements for a WOZ prototyping framework. Finally, a first prototype was developed and evaluated. Valuable results were produced and published in two conference papers (cf. [6] and [7]). In order to further proceed with the research the following steps were identified:

- Run several WOZ experiments in different experimental settings in order to understand the task of the wizard and find common problem patterns.
- Address these problems and define a modular wizard interface that optimally supports the wizard and allows for the testing of different application scenarios.
- Find a configurable software architecture that on the one hand supports the creation of wizard and client interfaces and on the other hand allows for a flexible integration of LTCs.
- Design and implement a WOZ prototyping framework based on the defined software architecture and evaluate its quality of supporting the wizard task.
- Evaluate the process of creating wizard and client interfaces for specific application scenarios and the efficiency of preparing, conducting and analyzing WOZ experiments. Finally, evaluate the diversity of application scenarios that are supported i.e. the amount of scenarios for which the framework can be used.

Expected Research Contribution

This research focuses on WOZ from the perspective of prototyping LTC-based software applications. As such it involves the analysis and comparison of existing prototyping tools, looks at their usability and assesses their support for different LTCs. A second aspect of the research looks more specifically at WOZ experiments in order to investigate the task of the wizard and the challenges a person in this role is facing. By doing so the goal is to identify supportive factors to be implemented in a novel WOZ framework that allows for prototyping, testing and evaluating software applications, which employ different kinds of LTCs.

In summary the overall contribution of the anticipated research outcome is to expand the body of knowledge on WOZ prototyping, specifically on the challenges of planning, designing and running WOZ studies, and to use this insight to create an application framework that offers an easy way of building WOZ prototypes that make use of different LTCs. Sub-contributions include the definition of a categorization model for WOZ experiments, the identification and classification of problems influencing a wizards performance, and a software architecture that allows for the flexible combination of differ LTCs.

References

1. Bradley, J., Mival, O., Benyon, D.: Wizard of Oz Experiments for Companions. In: Proceedings of the BCS Conference on Human-Computer Interaction, pp. 313–317 (2009)
2. Karpov, A., Ronzhin, A., Leontyeva, A.: A Semi-automatic Wizard of Oz Technique for Let's Fly Spoken Dialogue System. In: Sojka, P., Horák, A., Kopeček, I., Pala, K. (eds.) TSD 2008. LNCS (LNAI), vol. 5246, pp. 585–592. Springer, Heidelberg (2008)
3. Mäkelä, K., Salonen, E.-P., Turunen, M., Hakulinen, J., Raisamo, R.: Conducting a Wizard of Oz Experiment on a Ubiquitous Computing System Doorman. In: Proceedings of the IPNMD Workshop, pp. 115–119 (2001)
4. Rajman, M., Ailomaa, M., Lisowska, A., Melchiar, M., Armstron, S.: Extending the Wizard of Oz Methodology for Language-enabled Multimodal Systems. In: Proceedings of LREC, pp. 2531–2536 (2006)
5. Salber, D., Coutaz, J.: A Wizard of Oz Platform for the Study of Multimodal Systems. In: Proceedings of INTERACT and CHI, pp. 95–96 (1993)
6. Schlögl, S., Doherty, G., Karamanis, N., Luz, S.: WebWOZ: A Wizard of Oz Prototyping Framework. In: Proceedings of EICS 2010, pp. 109–114 (2010)
7. Schlögl, S., Doherty, G., Karamanis, N., Schneider, A., Luz, S.: Observing the Wizard: In Search of a Generic Interface for Wizard of Oz Studies. In: Proceedings of Irish HCI, pp. 43–50 (2010)

Time Affordances and Physical Mobility in the Context of Ubiquitous Technologies

Larissa Pschetz

University of Dundee, Perth Road, Dundee, DD1 4HT, UK
Edinburgh College of Art, Lauriston Place, Edinburgh, EH3 9DF, UK
pschetz@dundee.ac.uk, pschetz@eca.ac.uk

Abstract. This research aims to use design demonstrators to speculatively explore the relation between perceived time affordances and physical mobility in the context of ubiquitous technologies.

Keywords: user experience based approaches, speculative design,

1 Introduction

Recent interconnections of physical movement and communication systems are increasingly detaching geographic locations from underlying concepts of distance and co-presence [5]. As handheld devices allow a large number of activities to be performed on the go, individuals are encouraged to undertake longer hours of travel, and distance becomes ever more irrelevant. As a result, networks start to be structured more around individuals than around places, homes and contexts [6]. Daily schedules can be negotiated dynamically and become increasingly desynchronised from work, community and place, and hence from each other [15]. Technological developments not only contribute to free networks from geographical boundaries, but also set less utilitarian roles for spatial cognition [4] - as GPS-devices specifically and general automated travel technology enable people to navigate in completely unfamiliar settings with no prior planning [14].

Mobility thus is less constrained by geographic distances and is increasingly detached from spatial cues. It is however still and increasingly bound to individuals' schedules and personal relation to time [6]. Though there has been a lively debate among social scientists [17] on possible new notions of time provided by new technologies, there has been little research on relations between time perception and human mobility, and even less on the practical consequences of such notions for the design of new technologies. In HCI, supporting physical mobility is mostly associated to improving mobile infrastructure (e.g. accessibility [12], privacy [16] and services [13]), or to facilitating wayfinding e.g. [10]. While the first focuses on setting up the context of ubiquitous computing, the second approaches only one immediate aspect of physical mobility. As asserted by sociologist John Urry [1], demand for travel and movement tend to be treated as "either self-explanatory or exogenous to the system". Therefore, the present research intends to extend the current HCI discussion by

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 426–429, 2011.

considering physical mobility within a broader context. In particular, it will investigate the social ecologies around time perception and how it influences individual movements.

2 Research Hypothesis

The main hypothesis of this research is that physical mobility patterns in ubiquitous technological contexts are more strongly influenced by perceived time affordances than by geographic constraints. The idea is that, in highly technological contexts, individuals' movements are triggered, delayed and paced, not by notions of geographic distances nor by difficulties in localising routes, but by the perception that a schedule can be negotiated, meeting times can be squeezed, or that a place can be reached within a certain period of time. Technologies provoke multidimensional practices of time and new meanings of temporality [8], which consequently influence movement.

The challenge of this research is to harness technology that not only recognises, but celebrates the fact that perceived time affordances can vary greatly according to different social backgrounds and access to different technologies.

3 Methodology

The chosen methodology can be divided in four different phases as follows:

1) Development of demonstrators. Based on literature review and ethnography-based methods [11], major influences of current technological devices on physical mobility have been initially identified. By emphasizing time as the main driver for physical movement, mobility was then speculatively projected [9] into different scenarios, which did not intend to fit or present an accurate account of reality. Instead, it attempted to exaggerate critical aspects of pre-defined subjects in order to stimulate discussion. Such speculations were then translated into concrete products [18] that embodied these critical aspects in order to let participants experience alternative situations and later discuss their experiences. The design process consisted in exploratory and playful experimentations [7] with different kinds of hardware and information technology services, in which methods such as *experience prototyping* [3] and the already cited *scenarios* [2] were included. I will now present two examples of demonstrators developed in the context of this research.

a) The *Family Clock* speculates on the increasingly temporal flexibility provided by handheld devices. A family shares a clock whose rhythm is set collectively by its members according to their individual pace of life. The physical clock is placed in a common room, while its time is sped up or delayed throughout the day via family members' mobile phones until it reaches equilibrium. The clock allows families to choose their own temporal state, which can be slightly slower or faster than universal time.

b) The *Transportation Clocks* address issues of distance and social attachment. A pair of clocks constantly displays the time necessary to reach each other based on pubic transportation schedules. Serving as a dynamic souvenir of the person, its aim is not to fulfil utilitarian purposes, but to present a more personal way of perceiving time and distance, which in turn is expected to prompt discussion on the way these dimensions are generally perceived.

2) Experience evaluation. The developed design demonstrators will be presented to groups of users in order to discuss the relevance of each embodied subject. Though the circumstances of such assessment vary according to each demonstrator, participants will be asked to carry the artefacts or keep them at their homes for extended periods of time. During this period they will be asked to keep a written, visual or spoken diary of their impressions. Prompts developed based on ethnographic practices [11] will be provided to support the records. These practical experiments will be followed by focus groups or semi-structured interviews, which will allow us to understand the rationale behind methods adopted to record impressions, as well as formal features, potential for raising discussion, and the relevance of issues embodied by each designed artefact.

3) Refinements. It is expected that the evaluation phase will provide insights for refinements or for the development of new demonstrators, in an attempt to increase the complexity of proposed subjects. These refinements, in turn, will lead to new evaluations, in an iterative process.

4) Guidelines definition. Based on the insights gained at the evaluation phase, a set of guidelines will be defined to allow future researchers to either reproduce or apply the knowledge generated in the process.

4 Expected Contributions

This research will provide design-led evidence to support speculations on the future of physical mobility. By looking beyond the solving of problems related to mobility, towards an approach that explores the opportunities presented by existing and emerging technologies, the research expects to expand the scope of current HCI analysis on spatial mobility, generating new knowledge for mobility technology through: *a)* discursive accounts of user experiences of artefacts that play with perceptions of time and movement, *b)* a design based lens for looking deeply at mobility opportunities, *c)* a series of prototypes and artefacts that demonstrate and enable discussion on these opportunities, *d)* a set of design guidelines for implementing a future that has a rich and ubiquitous use of mobility technology.

Acknowledgement. This work was supported by Microsoft Research through its PhD Scholarship Programme.

References

1. Adey, P., Bissell, D.: Mobilities, meetings, and futures: an interview with John Urry. Environment and Planning D: Society and Space 28(1), 1–16 (2010)
2. Bødker, S.: Scenarios in user-centred design: setting the stage for reflection and action. Interacting with Computers 13(1), 61–75 (2000)

3. Buchenau, M., Suri, J.F.: Experience prototyping. In: Proceedings of the 3rd Conference on Designing Interactive Systems. ACM, New York (2000)
4. Burnettand, G.E., Lee, K.: The effect of vehicle navigation systems on the formation of cognitive maps. In: International Conference of Traffic and Transport Psychology (2005)
5. Cresswell, T.: On the Move, 1st edn. Routledge, New York (2006)
6. Elliott, A., Urry, J.: Mobile Lives, 1st edn. Routledge, New York (2010)
7. Gaver, B., Martin, H.: Alternatives: exploring information appliances through conceptual design proposals. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2000, pp. 209–216. ACM, New York (2000)
8. Hörning, K.H., Ahrens, D., Gerhard, A.: Do technologies have time? Time & Society 8(2-3), 293–308 (1999)
9. Kerridge, T.: Does speculative design contribute to public engagement of science and technology? In: Proceedings of Multiple Ways to Design Research (2009)
10. Leshed, G., Velden, T., Rieger, O., Kot, B., Sengers, P.: In-car gps navigation: engagement with and disengagement from the environment. In: SIGCHI Conference on Human Factors in Computing Systems, CHI 2008. ACM, New York (2008)
11. MacAulay, C.: Ethnography, theory and systems design: from intuition to insight. International Journal of Human-Computer Studies 53(1), 35–60 (2000)
12. Nilsson, V.: Ability by mobility. In: CHI 2001 Extended Abstracts on Human Factors in Computing Systems, CHI EA 2001. ACM, New York (2001)
13. O'Hara, K., Black, A., Lipson, M.: Everyday practices with mobile video telephony. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 2006, pp. 871–880. ACM, New York (2006)
14. Rajamani, R.: Vehicle Dynamics and Control, 1st edn. Springer, Heidelberg (2005)
15. Southerton, D., Shove, E., Warde, A.: Harried and hurried: time shortage and the coordination of everyday life. CRIC Discussion Paper, 47 (2002)
16. Tarasewich, P., Gong, J., Conlan, R.: Protecting private data in public. In: CHI 2006 Extended Abstracts on Human Factors in Computing Systems, CHI EA 2006, pp. 1409–1414. ACM, New York (2006)
17. Wajcman, J.: Life in the fast lane? towards a sociology of technology and time. The British Journal of Sociology 59(1), 59–77 (2008)
18. Zimmerman, J., Stolterman, E., Forlizzi, J.: An analysis and critique of research through design: towards a formalization of a research approach. In: Proceedings of the 8th ACM Conference on Designing Interactive Systems, DIS 2010, pp. 310–319. ACM, New York (2010)

Usability Evaluation in Software Development Practice*

Marta Kristin Larusdottir**

Reykjavik University, Menntavegur 1, 109 Reykjavik,
The Royal Institute of Technology,
Lindstedtsvägen 5, 10044 Stockholm
marta@ru.is

Brief description of the topic: Usability evaluation is explored by comparing the effectiveness of using different evaluation methods and by studying how usability evaluation is conducted by practitioners.

Research area. Usability, evaluation, agile development, Scrum, practice.

1 Description of the Background to the Research Topic

Usability evaluation is one of the core usability activities, defined in the ISO 9241:11 standard and in the ISO 13407 standard [5, 6]. During usability evaluation the evaluator actively defines the status of the usability of the system being evaluated. He or she often uses methods or techniques to focus the activities. Usability evaluation has proven to be a difficult task for the evaluators [9] so understanding how different the results are when various usability evaluation methods are used, understanding the context which evaluators work in and the hindrances they are experiencing is important.

Researchers have been interested in estimating the effectiveness of using various usability evaluation methods since early 90's see for example [2, 7, 8] to be able to advice software developers which method to use. The effectiveness of using a particular evaluation method has been measured by comparing the number of usability problems found by using various methods.

Comparing usability evaluation methods by counting usability problems that are found by each method was criticized by Wixon [15], were he states that research results on evaluation methods fails the practitioner, meaning people working actively on software development. Wixon [15] states that: "the development of real products is the only context sufficiently rich to produce the kind of nuanced examples that are needed to develop a differentiated and contextualized understanding of methods and techniques needed by practitioners". Some researchers have responded to his claim and have done studies that have the goal of supporting practitioners better, for example Uldall-Espersen and collages [14] where they study how useful usability evaluation results are to the practitioners.

* Short Paper for the Doctorial Consortium at INTERACT 2011.

** Supervised by Jan Gulliksen, The Royal Institute of Technology, Lindstedtsvägen 5, 10044 Stockholm, gulliksen@kth.se

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 430–433, 2011.

Agile processes for software engineering evolved to address perceived limitations of more established, plan-driven approaches for software development and have shown their worth in quickly developing reliable software [1]. The most widely used are XP and Scrum [11]. In Scrum the projects are split up in two to four weeks long iterations called sprints, each ending up with a potential shippable product. In Scrum self organizing and well compounded teams are heavily emphasized, typically with 6 – 8 interdisciplinary team members [12].

One of the benefits of using agile development processes was claimed to be that customers needs are taken more into account than when developing software using more traditional processes [12]. Traditional Scrum has been criticized for not involving real users in their software process and for not adequately address their usability needs, for example in [13]. Evaluation of the success of deploying user involvement methods in Scrum has so far only been anecdotal [11].

2 Description of the Research Topic and My Contribution

My first research question is: *Does the use of different usability evaluation methods give different results when evaluating the same software system?*

Through this question I have explored if the results of doing usability evaluation with users using the think aloud method and the results of using the inspection methods, heuristic evaluation and cognitive walkthrough are different. I have done two experimental studies on this subject, where the quantity of usability problems found by evaluators using these methods is compared [3, 4]. Furthermore, the quantity of serious problems found was studied. The reason for comparing these methods even though these could be used in different purposes is to understand how different the results are and what the strengths and weaknesses are of using these methods. That understanding will help practitioners to choose an evaluation method when evaluating a system they are developing and academics could use the results to be able to suggest new methods or techniques for evaluation [10].

My results from the first study show that the think-aloud method is the most effective method when comparing the results of using the three methods. Similar quantity of problems were found using the think-aloud method and using the heuristic evaluation, but a large extent of the problems found by using heuristic evaluation were false positive, meaning that these were not found during the think-aloud evaluation. In the second study one third of the problems found by using the think-aloud method were found by using the heuristic evaluation and again a fair amount of problems were false problems.

Through my second question*: How do practitioners integrate usability evaluation in the Scrum development process?* I explored how usability evaluation is conducted by software developers using the Scrum process in Iceland. By using a survey I studied how often developers evaluate usability, how that compares to other activities like testing and what the hindrances of doing usability evaluation are. The main findings show that unit, functional, system and acceptance testing are done to a wide extent [9]. Usability, security, performance, alpha and beta testing are much less emphasized. Interviews were conducted to exemplify how practitioners conduct usability testing and what they describe as the difference of usability and acceptance testing. Many of the respondents said that usability testing can only be done once or twice a year. The respondents did not have time for more frequent evaluation. Some examples from the interviews show that practitioners are willing to do formal usability testing on extensive parts of the system, but because the iterations in Scrum are short and the changes to the system in each iteration are small, formal usability testing does not fit into the project work.

Furthermore through my third research question: *What are the challenges for practitioners to integrate usability activities in Scrum development projects?* I have studied how other usability activities are integrated into the Scrum process to understand if the challenges of integrating usability evaluation are similar to the challenges integrating other usability activities. To do this we have conducted an interview study with 21 informants from Swedish software companies that use the Scrum process. The analysis of this study is still in progress, but the first results show that formal usability evaluation is not much emphasized. Instead more informal ways of evaluation are used, like observing users, getting feedback on the usability from users through group meeting, blogs or chats. Furthermore informal expert evaluations like peer reviewing are conducted by a number of the informants and evaluation with user representatives. The main obstacle for evaluating the usability formally with users more frequently than once a year is that the team does not have time. Non-functional requirements like usability are rarely stated or written. Some informants mention that it is hard to know when the usability of the system is good enough; it is always possible to extend the level of the usability. Furthermore the practitioners have different opinions on who is responsible for that the software system is usable, mentioning the whole team as being responsible, the project leader, the usability expert or actually no one as possible options.

To summarize, my research topic is usability evaluation, which is explored by comparing the effectiveness of using different evaluation methods and by studying how usability evaluation is conducted by practitioners using the software development process Scrum. This research topic is important, because even though usability evaluation methods are widely known in the research community of HCI, the methods are not widely used by the practitioners using the Scrum development process. Understanding how different the results of using various evaluation methods and how practitioners are evaluating their systems is of great importance for the HCI research community to be able to suggest new methods or ways to evaluate the usability of software systems that fit the practitioners.

References

1. Beck, K.: Extreme programming explained: Embrace change, 2nd edn. Addison Wesley, Reading (2004)
2. Desurvire, H.W., Kondziela, J.M., Atwood, M.E.: What is gained and lost when using evaluation methods other than empirical testing. In: Proc. People and Computers VII, pp. 89–102. Cambridge University Press, Cambridge (1993)
3. Frøkjær, E., Larusdottir, M.K.: Prediction of Usability: Comparing Method Combinations. In: Proc. of Idea Group Publishing Conf., Hershey, pp. 248–257 (1999)

4. Hvannberg, E.Þ., Law, E.L., Lárusdóttir, M.K.: Heuristic evaluation: Comparing ways of finding and reporting usability problems. Interacting with Computers 19(2), 225–240 (2007)
5. International Organization for Standardization, 1998: ISO 9241-11 Ergonomic requirements for office work with visual display terminals (VDTs) — Part 11: Guidance on usability, Geneva, Switzerland
6. International Organization for Standardization, 1999: ISO 13407 Human-centred design processes for interactive systems, Geneva, Switzerland
7. Jeffries, R., Miller, J.R., Wharton, C., Uyeda, K.: User interface evaluation in the real world: A comparison of four techniques. In: Proc. CHI 1991, pp. 119–124. ACM Press, New York (1991)
8. Karat, C., Campbell, R., Fiegel, T.: Comparison of empirical testing and walkthrough methods in user interface evaluation. In: Proc. CHI 1992, pp. 397–404. ACM Press, New York (1992)
9. Larusdottir, M.K., Bjarnadottir, E.R., Gulliksen, J.: The Focus of Usability Testing in the Software Industry. In: Proceedings of the World Computer Congress, Brisbane, Australia (2010)
10. Larusdottir, M.K.: Listen to your Users – The Effect of Usability Evaluation on the Software Development Process, Uppsala University (2009)
11. Salah, D., Petrie, H., Paige, R.F.: Towards a Framework for Integrating User Centered Design and Agile Software Development Processes. In: Proc. Irish CHI (2009)
12. Schwaber, K.: Scrum development process. In: OOPSLA 1995 Worshop on Business Object Design and Implementation (1995)
13. Singh, M.: U-SCRUM: An Agile Methodology for Promoting Usability. In: AGILE 2008, Conference, pp. 555–560 (2008)
14. Uldall-Espersen, T., Frøkjær, E., Hornbæk, K.: Tracing impact in a usability improvement process. Interacting with Computers 20, 48–63 (2008)
15. Wixon, D.: Evaluating usability methods: why the current literature fails the practitioner. Interactions 10(4), 28–34 (2003)

Website Customization: Exploring a Tag-Based Approach in the Australian Banking Context

Rajinesh Ravendran

Queensland University of Technology (QUT),
2 George Street, Brisbane, QLD 4000
{r.ravendran}@qut.edu.au

Description: Website customization is important to better fulfill the needs and wants of individual customers. Tags assigned to web resources are suitable to facilitate website customization.

Area: Website customization, online banking, tags.

1 Problem

This research addresses the problem of *website customization* in online banking, particularly in the Australian context. Australia has the fourth highest online banking adoption rate among developed nations (behind Canada, US and UK)[1] and Australian consumers prefer online banking over ATM, phone and branch[2]. The existence of large number of users who use online banking services on a regular basis presents a wide opportunity for customization. A recent study by Rahim and JieYing [11] on online banking customer satisfaction in the Australian context highlighted website customization as an imperative dimension, particularly among the younger generation, however, this dimension is poorly addressed.

This research is significant because website customization can enable online banking to be more responsive to the individual needs of each user [11]. Customization helps to bind customers into a long-term relationship despite a short-term discomfort [2]. Such relationships are vital as they directly translate into on-going profits [8]. In the banking context, customers are more likely to purchase a new product or service from their existing bank compared to a new bank [5]. For this reason, it is significant that online banking providers incorporate website customization as part of the offerings.

Prior research on website customization indicates that customization approaches can be grouped into two categories: static and dynamic [2]. Static approaches are typically user-based, where the user is responsible for initiating and carrying out the customization (e.g., content and feature selection during registration). Static approaches are simple, straight-forward and put users in control. However, they overlook the impacts of dynamic approaches such as the ability to predict what a user

[1] http://www.comscore.com/Press_Events/Press_Releases/2008/07/Canada_Online_Banking

[2] http://au.nielsen.com/news/20070426.shtml

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 434–437, 2011.

might find interesting or useful [2]. Also, users do not like to spend time configuring complex customization features [6]. Dynamic approaches, on the other hand, are system-based, with the system entirely responsible for initiating and carrying out customization (e.g., web usage analysis and collaborative filtering). They require minimal effort from users but are complex and sophisticated. Some of the known issues with dynamic approaches include expensive computational cost, technical issues and ethical concerns [9]. Additionally, dynamic approaches are likely to fail in dynamic settings [6].

This research proposes *tags* as a suitable technology to facilitate website customization. Tags, part of Web 2.0, represent Web resources (e.g., photo, video, people, etc) in ways familiar to individual users, primarily for personal information management (PIM). Tags are largely personal and contextual [6], and considered as a potential source of knowledge [1]. Recognized as an easy-to-use, dynamic and engaging technology, tags aid users to recall and retrieve information content [7] and when represented as tag clouds they facilitate visual information retrieval [3]. Also, the underlying meanings of tags may be discovered through semantic analysis to form associations between like-minded individuals [10]. These characteristics of tags make them suitable for customization and offer a valuable alternative approach.

In the financial space, tags assist personal financial management via tools such as Mint (http://www.mint.com) and Yodlee (http://www.yodlee.com), where a user can assign tags to transactional data for budgeting, expense tracking, etc. These tools, however, only allow tags to be assigned to financial transactions at a high level as category or description, but not at a lower level for details such as bank account or biller. This research explores the use of tags at a granular level in a broader context of personal financial management to facilitate customization in online banking. The types of customization proposed by Fung [2] (remembering, comprehension and associative) is used as a basis.

2 Claim

Tags are suitable to facilitate website customization alongside personal financial management. The proposed tag-based approach is user- and system-based with users as active participants of the customization process and the system providing customization based on user input (tags). The inclusion of both user and system in the customization process is advantageous for a complete customization. Such an approach will allow users to drive customization and simultaneously enable dynamic, system-based customization features.

Tags and tag clouds can offer a more intuitive and interactive website design. Since tags represent resources in a personal manner, tags can be used to personalize website interaction and comprehend user intentions via simple tag selections. Possible actions may be inferred on two tags based on their tag-resource association, which can enhance website interaction and experience. Also, based on semantic relations of tags across the community, relevant information can be aggregated and provided to users.

3 Methods

A mixed-method approach is proposed for this research, which includes a case study and software-based prototyping.

A case study has been conducted to identify the range of taggable resources found on online banking websites (mobile banking inclusive). The case study focused on two leading banks in Australia: Commonwealth Bank and Suncorp Bank. The personal banking websites of both banks were examined for potential taggable resources. Five resources were identified: account, reference, biller, application and message.

Software-based prototyping using an iterative approach is proposed to evaluate the suitability of tag-based website customization. This will allow users to provide feedback based on their 'real' experience that can be integrated into the prototype in the next iteration. A total of three iterations will be executed with each iteration comprising of three phases: design, development and evaluation. The prototype to be developed will be scenario-based where two key banking tasks, namely fund transfer and bill payment will be customized through tag integration. The prototype will be evaluated via an experimental design using pretest-posttest control group design [2]. A posttest questionnaire will be used to gather experiential feedback from participants on the utility and usability of the customization provided, as perceived by them.

The experiment results will be analyzed using ANOVA and interpreted by drawing on relevant literature in the space of HCI and ecommerce / banking customization. The results will be presented to the user-interaction and ecommerce research community.

4 Solution

The goal of the proposed tag-based approach is to simplify online banking activities, provide an intuitive interaction and offer personalized information to banking users. Online banking users will be able to quickly access and manage previous banking tasks (e.g., fund transfer, bill payment, etc) or resources (e.g., account, messages, etc) represented as tags; carry out banking activities by simply selecting appropriate tags (e.g., "Savings" + "Dad" to transfer money from savings account to dad's account); and discover related services and their popularity (usage) through associations with like-minded individuals based on tag similarity (e.g., "Vodafone" is semantically related to "Telstra" and "Optus", which are Australian telecom companies).

The outcome of this research will be a set of guidelines for designing and implementing tag-based customization. The guidelines will be akin to interaction design patterns in that offering solutions to different customization types and explaining when, how and why to use it (by quoting examples off the prototype).

5 Contribution

The contribution of this research will be the design and implementation of tags, expressed as interaction patterns, to facilitate website customization in online banking. This is expected to benefit financial institutions and system designers.

Acknowledgement. This research is sponsored by the Smart Services Cooperative Research Centre (CRC) of Australia (http://www.smartservicescrc.com.au).

References

1. Durao, F., Dolog, P.: A personalized tag-based recommendation in social web systems. Paper presented at the International Workshop on Adaptation and Personalization for Web 2.0 (2009)
2. Fung, T.: Banking with a personalized touch: Examining the impact of website customization on commitment. Electronic Commerce Research 9(4), 296–309 (2008)
3. Hassan-Montero, Y., Herrero-Solana, V.: Improving tag-clouds as visual information retrieval interfaces. Paper presented at the International Conference on Multidisciplinary Information Sciences and Technologies, InSciT 2006, Mérida, Spain (2006)
4. Hiltunen, M., Heng, L., Helgesen, L.: Personalized Electronic Banking Services. In: Karat, C.-M., Blom, J.O., Karat, J. (eds.) Designing Personalized User Experiences in eCommerce, pp. 119–140. Springer, Heidelberg (2004)
5. Marlow, C., Naaman, M., Boyd, D., Davis, M.: HT06, tagging paper, taxonomy, Flickr, academic article, to read. Paper presented at the 17th Conference on Hypertext and Hypermedia, Odense, Denmark (2006)
6. Nielsen, J.: Personalization is Over-Rated (1998), `http://www.useit.com/alertbox/981004.html`
7. Panke, S., Gaiser, B.: With my head up in the clouds: using social tagging to organize knowledge. Journal of Business and Technical Communication 23(3), 318–349 (2009)
8. Peppers, D., Rogers, M.: The one to one future: Building relationships on customer at a time. Currency Doubleday, New York (1996)
9. Pierrakos, D., Paliouras, G., Papatheodorou, C., Spyropoulos, C.: Web usage mining as a tool for personalization: A survey. User Modeling and User-Adapted Interaction 13(4), 311–372 (2003)
10. Qi Xin, M., Uddin, M.N., Geun-Sik, J.: The wordNet based semantic relationship between tags in folksonomies. Paper presented at the 2nd International Conference on Computer and Automation Engineering (ICCAE), Singapore (2010)
11. Rahim, M.M., JieYing, L.: An empirical assessment of customer satisfaction with Internet Banking applications: An Australian experience. Paper presented at the 12th International Conference on Computers and Information Technology, ICCIT (2009)

Acceptance and Speed of Animations in Business Software

Lisa Mattes, Martin Schrepp, Theo Held, and Patrick Fischer

SAP AG, Dietmar-Hopp-Allee 16, 69190 Walldorf, Germany
{Lisa.Mattes,Martin.Schrepp,Theo.Held,Patrick.Fischer}@sap.com

Abstract. Well-designed animations can improve the usability of graphical user interfaces for business software. It is crucial in this area that the user considers animations to be helpful and that they do not decrease perceived work efficiency. Thus, both the acceptance of animations and the selection of an appropriate animation speed are of high importance. We investigate those aspects in three explorative studies. Results show that animated interactions are considered to be useful by a considerable majority of participants. The observed settings for the animation speed show that quite fast animations are generally preferred. We demonstrate that for some of these animations the observed settings for animation speed can be explained by cognitive modeling.

1 Introduction

Well-designed animations can improve the usability of an interface. They can be used effectively to catch users' attention [1]. Animations can visualize changes on the user interface as a continuous transformation. This reallocates cognitive load to the visual system [2] and creates the impression of a direct manipulation of objects [3]. In addition, animations can improve the visual attractiveness of an interface [4].

Especially in the context of business software it is important that animations do not interrupt the work flow by being too fast or too slow. Thus, an appropriate setting for the speed of animations is very important in this domain. In addition, users should perceive animations as helpful in their daily work. Therefore, user acceptance is an important factor.

2 Investigations

We conducted three studies to investigate users' preferences concerning the speed of animations as well as the acceptance of the tested animations. In all three studies a quite similar procedure is used. First, the context of the current animation was explained to the participants. This should enable them to choose a suitable animation speed and to decide if the animation makes sense in this context. Second, the handling of the animation speed adjustment was explained by a short text and a video. After these instructions participants could switch to the adjustment of the animation speed. The upper part of the corresponding screen showed the initial state of the animation. Participants could start it by an instructed action (like clicking a button). Animation

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 438–441, 2011.

speed could be set by moving a slider on the bottom of the same page. Participants could adjust the speed and try out the animation as often as required to come to a decision. When the speed was considered to be adequate, the current setting could be confirmed by pressing a button. The initial speed of the animation was randomized per participant. Additionally, in studies 2 and 3 after confirming their setting, participants were asked if the animation was desirable in the described context (answer categories *Yes*, *No*, *Not Sure*). Here a lower animation speed implies a longer time until the animation ends. Thus, we report in the following the duration of an animation as a measure for the animation speed.

2.1 Study 1

This study scrutinizes an animation in the context of ecommerce. The animation is designed to provide feedback that a product chosen from the product catalogue in a web shop has been added to the shopping cart. After selecting the product, an overlay is opened that shows the current content of the cart, highlighting the new product with a green out-fading background.

Participants were recruited via email. The study was started by 52 persons. The data of 36 persons (average age: 34.42 years, 72% male, 22% female, 6% unknown) were complete and could be used in the analysis. In average the setting for the animation duration was 1838 ms (*SD* = 1492 ms). This observed value could be described by cognitive modeling. In the considered scenario the user had to move the visual focus to the shopping cart, perform a cognitive operation (‘Think’ [5]; ‘Mental Operator’ [6]) and had to process the product info in the cart (read 3 words). If we model this sequence, for example, with CogTool [5] this yields an estimated duration of 1850 ms, which is quite close to the observed value of 1838 ms.

2.2 Study 2

This study examines five different animations in a Customer Relationship Management (CRM) system:

- *Animations 1 - 3:* Participants should collapse three tables of differing sizes sequentially by clicking on an icon in the table header. Collapsing didn’t occur instantly, but progressed stepwise so that the table seemed to get squeezed. The number of rows of the table was varied to scrutinize the influence of this parameter on the preferred animation speed. The three tables contained 5, 15, and 25 rows respectively.
- *Animation 4:* This animation visualized hiding a block of information on a page. After clicking a corresponding icon in the header of the block, the block faded out until it completely disappeared.
- *Animation 5:* Lines in a table can be removed by clicking on a delete icon in the line. After this icon has been clicked, the content of the line first faded out completely and then the row collapsed until it wasn’t visible anymore. For this animation two sliders were offered to set the durations for the two components of the animation separately.

A link to the study was placed in a newsletter of a German university. The experiment was started by 60 persons. The data of 54 persons (average age: 22.58 years, 68.5% male, 29.6% female, 1.9% unknown) were complete and could be analyzed.

Table 1 shows the observed values for the animation duration and desirability ratings. For animation 5, the entry 5a refers to the time to fade out the content and 5b refers to the time for collapsing the line.

Table 1. Observed durations and evaluation of usefulness

Animation	Duration (ms) Mean (SD)	Animation desirable? (%)		
		Yes	No	NA
1	261 (199)	70.4	16.7	13.0
2	297 (173)	72.2	16.7	11.1
3	336 (314)	70.4	16.7	13.0
4	341 (223)	88.8	5.6	5.6
5 a	289 (222)	88.9	1.9	9.3
5 b	245 (192)			

The attitude towards the animations was very positive. All investigated animations were considered to be desirable by more than 70% of the participants, while durations of the animations were set to quite small values.

2.3 Study 3

This study investigates again an animation in the context of a CRM system. When users make errors in entering data, a corresponding error message is displayed at a specific location on the user interface. To increase salience of this message an animation is used. The error message first fades-in with a red background, then stays constant for a certain time and fades-out after that. Hence, this animation consists of three parts that participants can adjust with three separate sliders.

Participants were recruited via email. The study was started by 61 persons. The data of 51 persons (average age: 28.7 years, 77% male, 23% female) were complete and could be used in the analysis.

In total 60.8% of the participants considered this animation to be desirable in the described context, 17.6% found it not desirable, and 21.6% were unsure. Thus, also this animation is quite highly accepted. Concerning animation speed we observed the following mean values for the three components of the animation: Fade-in 652 ms (SD = 560 ms), staying constant 2672 ms (SD = 1013 ms), and fade-out 924 ms (SD = 607 ms). If we add the average durations of the single components we get a total of 4248 ms for the whole animation. Again this observed value for animation duration can be described by cognitive modeling. First, the user had to set the visual focus to the message and to read the message text (11 words). Under the assumption of a common reading speed of 300 words per minute [7] a cognitive model predicts 4750 ms to complete this cognitive sequence. Again this is close to the observed value.

3 Conclusion

A considerable majority of participants rated the investigated animations as desirable. But because of interindividual differences, differences in the acceptance of various animations, and general accessibility requirements there is a need of providing users with an option to disable animations in a software product.

There are substantial interindividual differences concerning the preferred animation speed as can be seen from the high standard deviations. This requires that users not only have to be allowed to disable animations, but also to adjust the speed to some degree. It is conceivable that there are differences because of users' age. However, an additional analysis didn't reveal such a result.

The investigated animations can be grouped into two categories. First, there are animations that visualize changes in the state of the user interface (expand/collapse areas, fade in/out content). Second, there are animations that guide the user's attention to a certain screen area and trigger a cognitive process (shopping cart, error message).

For animations of the first type, rather small durations are preferred (around 300 ms). For animations that trigger cognitive processes cognitive modeling helped predicting the animation durations quite accurately. This indicates that animations should run approximately as long as the cognitive processes triggered by the corresponding situations. Thus, we can hope to estimate the optimal duration of future animations that require thinking with the help of cognitive modeling.

References

[1] Hong, W., Thong, J., Tam, K.Y.: Does Animation Attract Online Users' Attention? The Effects of Flash on Information Search Performance and Perceptions. Information Systems Research 15(1), 60–86 (2004)

[2] Robertson, G.G., Mackinlay, J.D., Card, S.K.: Cone Trees: Animated 3D Visualizations of Hierarchical Information. In: Proc. CHI 1991, pp. 189–194. ACM Press, New York (1991)

[3] Thomas, B.H., Calder, P.: Applying Cartoon Animation Techniques to Graphical User Interfaces. ACM Transactions on Computer-Human Interaction 8(3), 198–222 (2001)

[4] Sundar, S.S., Kalyanaraman, S.: Arousal, Memory, and Impression-Formation Effects on Animation Speed in Advertising. Journal of Advertising 33(1), 7–17 (2004)

[5] John, B., Prevas, K., Salvucci, D., Koedinger, K.: Predictive Human Performance Modeling Made Easy. In: Proc. CHI 2004. ACM Press, New York (2004)

[6] Card, S.K., Moran, T.P., Newell, A.: The keystroke-level model for user performance time with interactive systems. Commun. ACM (1980)

[7] Fry, E.B.: Teaching faster reading: a manual. Cambridge University Press, Cambridge (1963)

Developing Mobile Remote Collaboration Systems for Industrial Use: Some Design Challenges

Leila Alem and Weidong Huang

CSIRO ICT Centre,
PO Box 76,
Epping NSW 1710 Australia
{Leila.Alem,Tony.Huang}@csiro.au

Abstract. Many real-world collaboration scenarios involve a helper remotely guiding a worker performing a task requiring the manipulation of physical objects/devices. Systems currently available for remote guiding have limitations for industrial use, particularly in terms of supporting the mobile aspect of work. To meet these needs and as part of our work for the mining industry, we have developed two systems. These systems use a wearable computer and a wearable display to support the mobile aspect of the remote collaboration. In this paper, we review existing work in remote guiding and state their limitations in relation to industrial needs. We then present two mobile remote collaboration systems that we have designed and developed to support the maintenance of mining equipments. Designing for real world use is challenging, systems need to be easy to use and be able to operate in various environmental conditions. Systems also need to support the mobility aspect of work and support different roles of the collaborators.

Keywords: Remote collaboration, mobile collaboration, remote gestures, augmented reality, usability.

1 Introduction

There are a range of real world situations in which remote expert guidance is required for a local novice to complete physical tasks. In maintenance, for example, an expert may be guiding remotely a technician into repairing a piece of equipment. There is a growing need for systems or platforms that support collaboration between a worker and a helper, ensuring the effective delivery, just in time, of remote guidance.

Prior research on remote collaboration has demonstrated that providing access to a shared visual space and supporting the richness of hand gestures are critical to the success of remote guiding [2, 6, 7, 9]. However, existing systems either confine collaborators in fixed desktop settings (e.g., [2, 9]), or support only the pointing gesture (e.g., [8]). This limits their use in real world applications. For example, in the mining industry, when a sophisticated machine is not operational, onsite maintenance technicians require urgent input from a remote specialist/expert, as the time lost in the machine not being used translates into a loss in productivity. In addition, on-site technicians are highly mobile workers; in order to repair a machine, they are often

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 442–445, 2011.

required to go to the location of the machine, move around the machine, fetch tools and inspect equipment. This in turn requires that the systems be wearable and wireless, and that the systems function in various working conditions.

To meet these mobility requirements, we developed two systems: HandsOnVideo for a fixed remote helper guiding a mobile on-site worker [1, 4] and HandsInAir for a mobile remote helper guiding a mobile on-site worker [3]. These two systems were developed as part of our Human System Integration project within the CSIRO's Minerals Down Under (MDU), a National Research Flagship Program.

2 HandsOnVideo

Workers in mine sites wear helmets for safety. HandsOnVideo uses the helmet as a support for mounting a camera and a near eye display (see the left image in Figure 1). This setup supports the mobility of the worker. It is also hands free, hence allowing the worker to use his/her hands for maintenance operations. The near eye display is a small device with two small screens at the top of the worker field of view. By looking up, the worker can see what is displayed on a virtual screen, that is, the video of the workspace overlaid with the helper gestures (pointing to objects and showing orientation). With this setup, the worker is also able to be aware of his/her workspace and the surrounding environment.

The helper station is a large touch-display table located in an office or a workshop. It consists of three components: 1) a shared visual space that displays the video of what the worker sees (captured by the camera on the helmet of the worker); 2) a panoramic view of the worker's workspace, which the helper can use to maintain situation awareness; and 3) four storage areas on both sides of the shared visual space, which allow the helper to store specific scenes of the workspace for future use.

The right image in Figure 1 illustrates how the system works. First, the video of the worker's camera is sent to the helper side and displayed on the shared visual space (arrow 1) of the touch table. The helper uses his/her hands over the shared visual space to show/explain how a procedure should take place. The helper hand gestures are captured by a camera mounted on top of the display (arrow 2). The captured hand gestures are then sent to the worker side and displayed on the near eye display (arrow 3).

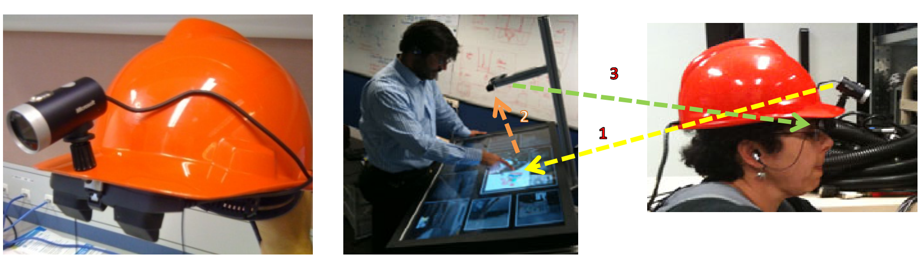


Fig. 1. Left: helmet with camera and near-eye display. Right: system work flow

3 HandsInAir

Expert helpers may be out of the office and on the move. HandsInAir was developed to explore the feasibility of supporting the mobility of the helper while using an interface/system similar to the one worn by the worker. The collaborators have the same hardware configuration as shown in the left image in Figure 1. This configuration not only allows the helper to move around, but also allows him/her to point to remote objects and use more complex hand gestures on a virtual display of the workspace, removing the need for a large physical display.

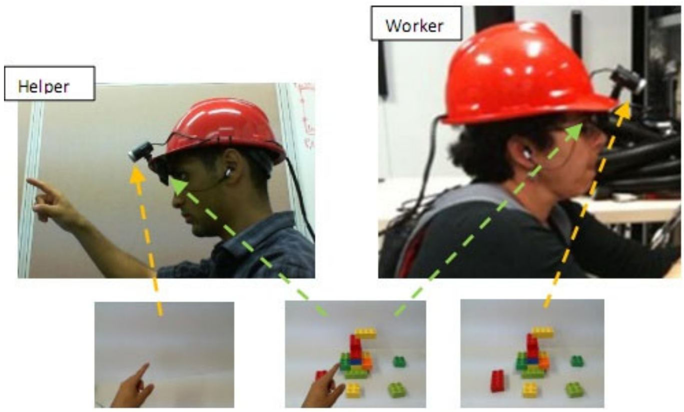


Fig. 2. Camera captures and hand gestures showed on the near eye displays

Figure 2 illustrates how the system works. Once a connection between the helper and the worker is established, the system initializes two video streams between the two sides. First the video of the worker camera is sent to the helper side and displayed on the near eye display. The helper sees the workspace of the worker and uses his/her hands to support instructions. The helper's hand gestures are captured by the helper camera and overlaid on the local videos of both the worker and helper (see Figure 2). Initial testing of HandsInAir suggests that it is easy for the users to point to remote objects using the virtual display, and that it is easy and natural to perform gestures on the virtual display in order to show, for example, how two pieces of Lego toys should be assembled in a specific way.

4 Designing for Real World Use: Some Issues

Designing for real world use is challenging, systems not only need to be easy to use, but also need to support the mobile and the flexible aspect of work.

Supporting mobility of the helper comes with a price. The helper unit in HandsOnVideo provides a panoramic view of the workspace of the worker; this allows the helper to point to the area of attention outside the field of view of the worker. However, this awareness support is no longer available in the current setting of HandsInAir. Research is needed to explore how awareness of the worker's environment should be provided to the helper.

Supporting the flexibility of work means that the systems for real world use should be able to operate in various environmental conditions, specifically lighting conditions. Previous work has demonstrated the value of unmediated gestures. For example, the projection of hand gestures into the workspace may lead to a gain in task performance [5, 6]. The display of the workspace with an overlay of the hand gestures may lead to an increase in satisfaction with the collaborative effort [9]. We investigated the option of projecting the hands using Augmented Reality as well as the option of using a see through system. The dependence on the right lighting conditions for the projection option and the limitation of the field of view for the see through system resulted in disregarding these options.

Both HandsOnVideo and HandsInAir make use of a near eye display. The use of the near eye display allows these systems to operate in various lighting conditions. However, this comes at a price. In both systems, the hand gestures are displayed on a small virtual screen, in which the workspace plus the hands of both the worker and the helper are displayed. Our users playing the role of the worker reported that sometimes they had to remove their hands from the workspace in order to leave some space for the hands of the helper.

In summary, we have presented two systems that we have developed to support the mobility aspect of work. These two systems draw on the latest research in remote guiding systems. While they are designed to be suitable for industrial use, some research effort still remains for further improvements and refinements.

References

1. Alem, L., Tecchia, F., Huang, W.: HandsOnVideo: Towards a Gesture based Mobile AR System for Remote Collaboration. In: Alem, L., Huang, W. (eds.) Recent Trends of Mobile Collaborative Augmented Reality Systems, pp. 127–138 (2011)
2. Fussell, S.R., Setlock, L.D., Yang, J., Ou, J., Mauer, E., Kramer, A.D.I.: Gestures over video streams to support remote collaboration on physical tasks. Human-Computer Interaction 19, 273–309 (2004)
3. Huang, W., Alem, L.: HandsInAir: A Wearable System for Remote Collaboration. In: APWCS 2011, pp. 171–174 (2011)
4. Huang, W., Alem, L.: Supporting Hand Gestures in Mobile Remote Collaboration: A Usability Evaluation. In: BCS HC 2011 (2011)
5. Kirk, D.S., Rodden, T., Stanton Fraser, D.: Turn It This Way: Grounding Collaborative Action with Remote Gestures. In: CHI 2007, pp. 1039–1048 (2007)
6. Kirk, D.S., Stanton Fraser, D.: Comparing Remote Gesture Technologies for Supporting Collaborative Physical Tasks. In: CHI 2006, pp. 1191–1200 (2006)
7. Kraut, R.E., Fussell, S.R., Siegel, J.: Visual information as a conversational resource in collaborative physical tasks. Human-Computer Interaction 18, 13–49 (2003)
8. Sakata, N., Kurata, T., Kato, T., Kourogi, M., Kuzuoka, H.: WACL: supporting telecommunications using wearable active camera with laser pointer. In: Fensel, D., Sycara, K., Mylopoulos, J. (eds.) ISWC 2003. LNCS, vol. 2870, pp. 53–56. Springer, Heidelberg (2003)
9. Alem, L., Li, J. (2011) A Study of gestures in a video mediated collaborative assembly task. International Journal on Advances in Human Computer Interaction, Article ID 987830, 7 (2011), doi:10.1155/2011/987830

Experiences of Online Co-creation with End Users of Cloud Services

Kaarina Karppinen[1], Kaisa Koskela[1], Camilla Magnusson[2], and Ville Nore[2]

[1] VTT Technical Research Centre of Finland, Kaitoväylä 1, 90570 Oulu, Finland
{kaarina.karppinen,kaisa.koskela}@vtt.fi
[2] F-Secure Ltd. Tammasaarenkatu 7, 00180 Helsinki, Finland
{camilla.magnusson,ville.nore}@f-secure.com

Abstract. This paper describes an online co-creation study done via an online co-creation platform Owela as well as shares industrial experiences and lessons learnt about the study. The Owela study was conducted in order to provide a deeper understanding of users' perceptions of cloud services and their security. By utilising the online co-creation platform it was possible to get quick and easy contact to geographically distributed cloud service users. For the company Owela offered an efficient way to apply online user participation while for the end users Owela enabled convenient participation in various co-creation activities regardless of time and place. As an end result of the study the voice of the cloud service users was turned into several new business ideas.

Keywords: Online co-creation, Owela, Cloud services, End users.

1 Introduction

Cloud phenomenon is one of the main global trends washing through the ICT industry [1]. Today the cloud is also reality in consumers' life as typical web users use cloud services daily whether they know it or not [2]. This novel way of offering software, platforms and infrastructure as a service via the internet opens up new possibilities and advantages for companies such as F-Secure, a provider of consumer cloud security and storage solutions.

In order to create new business opportunities within the cloud security domain there is a need to understand what the user perception of cloud services and their security is, and how it is formed. However, reaching geographically distributed cloud service end users via traditional user involvement methods is challenging. To overcome this challenge, we used an approach where co-creation with cloud service end users was done via an online co-creation platform Owela (http://owela.vtt.fi/owela/introduction). Owela is an online living lab that builds on social media features for co-creation activities and open innovation [3]. It was launched in April 2007 and is administrated by VTT Technical Research Centre of Finland. The conducted four week-long Owela study involved a use of various online co-creation tools enabled by the platform such as discussions, commenting, blog posting, idea posting, and chat sessions. In this paper we are presenting how the study was conducted as well as sharing experiences and lessons learnt from it.

P. Campos et al. (Eds.): INTERACT 2011, Part I, LNCS 6946, pp. 446–449, 2011.

2 Background

Advances in information and communication technologies give companies new opportunities to include users to their processes [3]. According to Swahney et al. [4] online based virtual environments allow a company to engage a larger number of end users without significant compromises on the richness of the interaction. Virtual environment also increase the speed and persistence of user engagement and the episodic and one way interaction with end users can be transformed into a continuous dialogue [4]. Through online collaboration users can also be involved in co-creation activities regardless of time and place.

3 Description of the Owela Study

The purpose of the Owela study was to discuss with the cloud service users about how they actually use the services and how the existing services could be improved as well as to co-create ideas for new services. The participants were selected based on their responses to an online survey carried out in June 2010 [2]. A total of 128 end users were invited to join the Owela workspace. From them 47 registered within one week from the invitation. Also 4 representatives from F-Secure participated actively in Owela during the whole study. The company representatives were separately identified so that the users were all the time aware which actions were made by other users and which by the company representatives. In addition, two VTT researchers facilitated the online co-creation. Table 1 gives more information about the study.

Table 1. Details of the Owela study

Length	4 weeks (24.9.-22.10.2010)
Participants	47 end users 4 company representatives 2 researchers
Tools for communication	Discussions (320 postings) Blogs (60 comments) Idea posting and commenting (14 ideas, 80 comments) Chat sessions (5 x 2 hour sessions)
Schedule for participation	Schedule for participation was optional (users could participate anytime) except for the chat sessions for which the time was announced beforehand
Motivation and rewarding	Activity points and rewards Active participation of F-Secure and VTT representatives Weekly email reminders

In the Owela workspace designed for the study, several specific features were used to achieve continuous communication between the end users and company representatives. First of all, Owela enabled active online discussion. The discussions revolved around predefined theme questions which were given to the participants each week. The participants answered these questions and were able to comment each other's answers thus creating discussion. Also F-Secure's and VTT's representatives were actively involved in the discussions to steer them and to probe emerging interesting topics. Secondly, in the Owela blog section users could write down their daily experiences with usage of cloud services according to each week's theme. The participants were encouraged to e.g. share negative and positive experiences with cloud services, compare different services, consider their network identity, and talk about situations that made them feel insecure. Thirdly, the participants had the chance of posting ideas about new cloud services. When posting an idea the participants were asked to write down both the idea that they had in mind and the problem that the idea would solve. This way also other participants had the opportunity to give solution ideas for the stated problems. Finally, five two-hour chat sessions were organized. The purpose of the chat sessions was to promote the direct and synchronized communication between the participants and the company representatives.

Activation and motivation of participants was considered very important in order to receive as good research results as possible. The primary way of activation was the active company representative participation in the discussions by asking questions and commenting the participants' postings. Another way of activation was to send weekly emails to the participants. The emails included current issues such as the theme and tasks for the week as well as a reminder of the chat session schedule and collecting the activity points. The third way to activate the participants was rewarding. The participants earned activity points for their actions at the Owela workspace and these points could be used as lottery tickets in a weekly movie ticket lottery as well as for earning a product gift in the end of the study.

4 Experiences and Lessons Learned

From F-Secure's point of view, online co-creation was seen to be a fruitful new way of interacting with end users. Compared to traditional end user research methods such as focus groups, interviews and usability testing used by F-Secure, this method turned out to be much more flexible, which was a benefit to both the end users and the company. It allowed participating end users to give their input whenever they had a few minutes of extra time, without having to leave their home or office. For F-Secure, it enabled defining various research goals during the study, reacting to feedback, and modifying the goals accordingly. Moving focus to new topics that may emerge from the user community during the study is usually hard to do with traditional, often quite rigid forms of end user research. Online co-creation thus enabled a continuous and rich dialogue between the company and end users.

However, it is important to note that Owela or any other online co-creation tool alone cannot ensure the success of a consumer research project. In order to reach the goals, commitment from the company as well as skilful facilitators, for keeping the discussions going and stimulating participants, were essential. Key factors for achieving success were: 1. active participation of all the parties in online conversations, 2. maintaining an active and encouraging atmosphere throughout the whole duration of the study, which motivated users to login frequently, and 3. enabling users to freely share their creative thoughts without the fear of ridicule or pressure to conform.

For F-Secure, the study produced immediately actionable results. It provided valuable insights into consumers' usage and perceptions of cloud services. It also generated 14 new consumer product ideas, which were entered into F-Secure's idea funnel for further development and evaluation. Some examples of the general themes of the product ideas are 1) providing a more secure online shopping and browsing experience, 2) allowing users to control their privacy online, and 3) securing users' cloud based data in various ways. Furthermore, the study provided a way for F-Secure to obtain feedback on upcoming products from mainstream end users, who may not have enough interest in technology to take part in traditional beta testing communities or company-specific, often quite technical, discussion forums.

5 Conclusions

The online co-creation study conducted via Owela provided F-Secure with a deeper understanding of users' perceptions of cloud services and their security. By utilising the online co-creation space it was possible to get quick and easy contact to geographically distributed cloud service users. Owela acted as a continuous communication channel between the company representatives and the cloud service users. Active online participation by all parties and energetic and encouraging atmosphere created in Owela were the key factors for achieving successful results. For F-Secure Owela offered an efficient way to apply online user participation while for the users Owela enabled participation in various co-creation activities anytime anywhere. The voice of the cloud service users was turned into several new business ideas that are further explored within the company.

References

1. Dikaiakos, M.D., Katsaros, D., Mehra, P., Pallis, G., Vakali, A.: Cloud Computing: Distributed Internet Computing for IT and Scientific Research. IEEE Internet Computing 13(5), 10–13 (2009), doi:10.1109/MIC.2009.103
2. Karppinen, K., Koskela, K., Koivumäki, T.: Usage of Cloud Services - A Survey for Consumers in USA, Japan and Finland. In: 1st International Conference on Cloud Computing and Services Science - CLOSER 2011, Noordwijkerhout, The Netherlands, May 7-9, pp. 305–308 (2011)
3. Näkki, P., Antikainen, M.: Online Tools for Co-design: User Involvement through the Innovation Process. In: The Workshop: New Approaches to Requirements Elicitation & How Can HCI Improve Social Media Development?, NordiCHI, Lund, Sweden, October 18-22 (2008)
4. Di Gangi, P.M., Wasko, M.M., Hooker, R.E.: Getting Customers' Ideas to Work for You: Learning from Dell How to Succeed with Online User Innovation Communities. MIS Quarterly Executive 9(4), 213–228 (2010)
5. Swahney, M., Verona, G., Prandelli, E.: Collaborating to Create: the Internet as a Platform for Customer Engagement in Product Innovation. Journal of Interactive Marketing 19(4), 4–17 (2005)

Interactive Installations: Tales from the Trenches

Pedro Campos[1,2], Miguel Campos[3], and Joaquim A. Jorge[2]

[1] University of Madeira, Campus Universitário da Penteada, 9000-390 Funchal, Portugal
[2] VIMMI Group, INESC-ID Lisbon, R. Alves Redol, 9, 1000-029 Lisbon, Portugal
[3] WowSystems, Lda., R. Mary Jane Wilson, 21 G, 9050-446 Funchal, Portugal
pcampos@uma.pt, miguel.campos@wowsystems.pt, jaj@inesc-id.pt

Abstract. Breakthrough innovation can be interpreted as research translated into products that the market accepts. The process of market translation of several products developed by WowSystems, a Portuguese company specialized in novel interaction paradigms, is explained in this paper as a case study shedding some light into how innovation centers can better promote innovation, in the form of well-succeeded products. We describe two paradigmatic "tales from the trenches" and conclude with some guidelines that were outlined on the basis of more than three years delivering fifty interactive installations.

Keywords: Interactive Installations, Guidelines, Case Study, User Interfaces.

1 Introduction

Designing interactive installations for diverse venues and different contexts has become increasingly popular [1]. Science centers wish to exploit the interactive, often surprising, element to bring more visitors and to explain difficult scientific concepts in a more appealing way. Museums wish to attract visitors from all ages and promote collaborations between them. Retail stores and shops have also started to embrace interactive installations as a way to improve their relationship with existing clients as well as capture the attention and interest of new segments, exploiting installations featuring the so-called "wow!" effect. Moreover, the speedy evolution in computing power available, as well as the decreasing cost in display technologies, such as projectors and LCD displays, has also led to an increased level of interest from retailers wishing to improve their stores' attractiveness, museum curators wishing they had a nicer way to display the rich-ness of cultural heritage, science centers' managers who are simply technology enthusiasts and thrive with the idea of refurnishing their centers with the latest innovations. And the list goes on and on. Success seems almost certain in such a context. But sometimes the client says the project fails to deliver the intended effect. So, what went wrong?

This question is timely and, therefore, very relevant: what risks and opportunities do designers and developers face when delivering real world innovative interactive installations? By defining *breakthrough innovation* as research prototypes translated into products that the market accepts, we describe our own experience in bridging research prototypes developed at the University of Madeira and at INESC-ID Lisbon, and studying how they were translated into real world applications in different

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 450–453, 2011.

contexts. We have been lucky enough to work around several practitioners' issues and risky situations in this field, and we have been working towards compiling sets of guidelines based on both successful and not so successful projects. While some of the more than fifty interactive installations already deployed were solely created as experiential activities, providing an increase in the level of learning by adding facts to an already well-formed conceptual model, others were designed to enact a reflective activity, thus supporting a restructuring learning where new conceptual frameworks need to be built. Based on this experience, we have summarized into a set of guidelines some ways to help interaction designers survive and perform well when the expectations are increasingly getting higher.

2 From the Interactive Installations' Trenches

It has been argued that the identification of breakthrough ideas at the very forefront of the innovation process is a key factor towards the creation of substantial innovation [2]. However, the managerial process of breakthrough innovations, as well as their inhibitive factors, remains far from being understood [2].

In this paper, we briefly describe our own experience with WowSystems, a Portuguese company specialized in new digital media, novel interaction paradigms and interactive installations. Because of the very nature of its core business, innovation is – and should continue to be – a main concern of the company. Founded in early 2008, WowSystems' main focus has been to professionally create useful and usable interfaces that make people *go "Wow!"*.

Two tales from the trenches we selected are a cultural interactive exhibition and an interactive shoe store. WowSystems designed a set of sensor-based installations in a cultural exhibition organized by the Direction of Cultural Affairs, which aimed at showing the visitor the cultural richness that formed the streets of Funchal (Portugal). The concepts of the exhibition revolved around promoting awareness about, and foster a better understanding of, the cultural tourism that can be performed by simply walking through strategic streets and watching certain buildings, sites, and heritage. To better complement the exhibition's traditional large-format printed panels, the organizers wanted to have the interactive factor as a means to add value to the visitor's experience. The left photo in Fig. 1 illustrates an installation in this exhibition.

In a similar project, we designed and installed and interactive mirror for a shoe shop, illustrated in the right photo in Fig. 1. The client's expectations included: That the shoe shoppers would step inside an RFID-tagged shoe and watch themselves inserted into a real time virtual scenery related to the type of shoe they were trying on. Our design had the following characteristics: As a shopper walks around the experimenting floor, the shoe's RFID tag is read by the reader, then the model's attributes are fetched from the product database, sent to the multimedia server which displays two synchronized scenarios: one for two top-down projections (left photo) and one for the front, "mirror-like" view (right photo).

The "mirror-like" front view displays the shopper in real time and places her on a virtual scenery by using a motion detection and silhouette extraction algorithm. This algorithm is adaptive regarding the different lighting conditions at the shop – usually brighter during the day and darker at dusk and night. Top-down projections feature views of the streets or sidewalks that are typical of the city that the virtual scenery replicates. For instance, the photo shows a shopper trying a shoe model that had a design inspired by modern life in Tokyo. Therefore, our interactive mirror displays a scenery based around Tokyo's neon signs and bright buildings. Simultaneously, the floor projections display a Tokyo sidewalk with Japanese signs and warnings, as well as other visual elements, and add interactivity by displaying neon lights over the floor according to the shopper's position.

Fig. 1. Two examples taken from the trenches: an interactive cultural exhibition (left) and a virtual mirror for a shoe store

The first interactive exhibition project went very well and post-project analysis suggests that good communication with museum curators, artists and designers was crucial to its success. A good mapping of the interaction styles into the installations and the exhibitions' theme was also important to grab the visitors' attention.

Upon the final installation of the shoe store, however, the solution didn't fulfill the client's expectations. Post-project analysis suggests one of the reasons this happened was simply because the expectations were put too high. Contrarily to other projects delivered to this client by other companies, however, our solution fitted the consumers' profiles very well, and the high satisfaction levels that shoe shoppers expressed were crucial in helping us defend the project's solutions.

3 Conclusions

Because of today's diversity of possible technological combinations for any interactive installation, the solution space has become too large. This, we argue, is an issue that contributes to increasing risks in interactive installations' development. And it's one of the reasons why it is surprisingly easy to create bad designs.

Secondly, experience has shown that during an economic crisis some clients start focusing on finding excuses for not admitting a project's success—and therefore not paying. The problem with frontline interaction design is that it's fairly easy to debate or discuss the final results of an installation: People's tastes are highly subjective and vary a lot. Requirements engineering as a discipline has many principles, techniques, and methods devoted to traditional software development. However, in terms of validating interaction design requirements, research literature is somewhat scarce. More effort should be put into how we can more effectively work collaboratively with stakeholders in order to better define the interaction design aspects of any given project's requirements.

The guidelines we identify on the basis of our experience are our own way to deal with the difficulties, risks and opportunities that come up in this field. They could prove useful for other interaction designers, business managers, and even clients, as a way to design, develop and install better products. These guidelines include, but are not limited, to the following.

Making the Vision Stand Out. It is a good idea to hang exhibition posters that feature interactive installations, photos of the visitors, and, for instance, give away free tickets, whenever applicable, so that engineers and designers can experience the installations the exact same way clients and users do.

Know the customer from the client. Interactive installations are meant to be fun, enriching, and enticing to everyday customers. A successful installation will attract more customers and more business, therefore making your client happy. The focus should be on your client's customers and not on your clients. A good defense mechanism to support design decisions is to convincingly and accurately document the customers' satisfaction and deliver that documentation to your client with a partnership attitude. Collecting evidence such as happy customers' photos, videos of people interacting with the installations, even surveys or informal interviews, can be useful to convince your client, especially if cross-checked with sales or other business figures. Please your client's clients.

Carefully manage client expectations. One way to achieve this is to present the client with architectural designs of how the interactive installation will look at the end of the project. If we provide the client with a visual scale and 3-D layout, the idea can be conveyed in a way that gives all stakeholders a feel of how the physical space will be used for the installations, just like in architectural programs.

Acknowledgements. This work has been supported by FCT (the Portuguese Foundation for Science and Technology) through grant PTDC/EIA-EIA/116070/2009.

References

1. Liu, W., Fernando, O.N., Cheok, A.D., Wijesena, J.P., Tan, R.T.: Science Museum Mixed Reality Digital Media Exhibitions for Children. In: Proceedings of the Second Workshop on Digital Media and Its Application in Museum & Heritage, December 10-12, pp. 389–394. IEEE Computer Society, Washington, DC (2007)
2. Funnel or tunnel? A tough journey for breakthrough innovations. In: Proceedings of the 4th IEEE International Conference on Management of Innovation and Technology, ICMIT, pp. 368–373 (2008)

A Conceptual Framework for Modeling Awareness Mechanisms in Collaborative Systems

Fernando Gallego, Ana Isabel Molina, Jesús Gallardo, and Crescencio Bravo

Computer Science Department. University of Castilla - La Mancha
Paseo de la Universidad, 4, Ciudad Real, Spain
{Fernando.Gallego,AnaIsabel.Molina,Jesus.Gallardo,
Crescencio.Bravo}@uclm.es

Abstract. Awareness is defined as the perception of the activities being carried out by members of a team in a given context. The use of awareness support techniques reduces the effort needed to realize communicative and collaborative tasks. Therefore, this leads to an increase in group work productivity. In this article we propose a conceptual framework which improves upon the process of building interactive collaborative systems as proposed by CIAM (a methodological framework for designing interactive applications for group work) by taking the dimension of awareness into consideration.

Keywords: Awareness, Meta-modeling, Groupware.

1 Introduction

The use of technologies for awareness [1] support in collaborative systems allows users to stay informed about the activities that are being carried out by other users, as well as in which parts of the shared working area they are working and how those activities are being carried out. Therefore, the incorporation of elements for awareness support makes it possible to reduce the meta-communicative efforts which are needed to carry out collaborative activities oriented towards promoting real collaboration between group work members. Other proposals [2, 3, 4] in the field of awareness modeling in collaborative systems usually focus on the domain of awareness design in interactive systems, but do not take into account any specific technological implementation.

Molina et al. [5] have reviewed the different existing proposals in the field of model-driven specification of collaborative interactive systems. We can conclude that none of them provides a proper specification for users' needs regarding awareness much less any systematic or computational support for modeling this characteristic. One of these proposals is the CIAM methodological framework [6]. Its related notation, called CIAN [7], provides support for modeling collaboration, interaction and shared context. Shared context is defined as a set of objects where both the objects and the actions performed on them are visible to a set of users [8]. In group work tasks with a shared context (collaborative tasks), awareness mechanisms are very important because they provide a noticeable improvement to group work productivity. Therefore, the main objective of our work is: *to incorporate the*

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 454–457, 2011.

modeling of these awareness aspects into the modeling of collaborative tasks in the context of CIAM.

In the next section we briefly show our framework for modeling awareness mechanisms, and in Section 3 we present the conclusions and future lines of work derived from this work.

2 A Conceptual Framework for Designing Awareness Support in Collaborative Tasks

To specify which group awareness concepts should be included in the CIAN notation we started by defining a meta-model (Figure 1) which includes the concepts and groups them into different views:

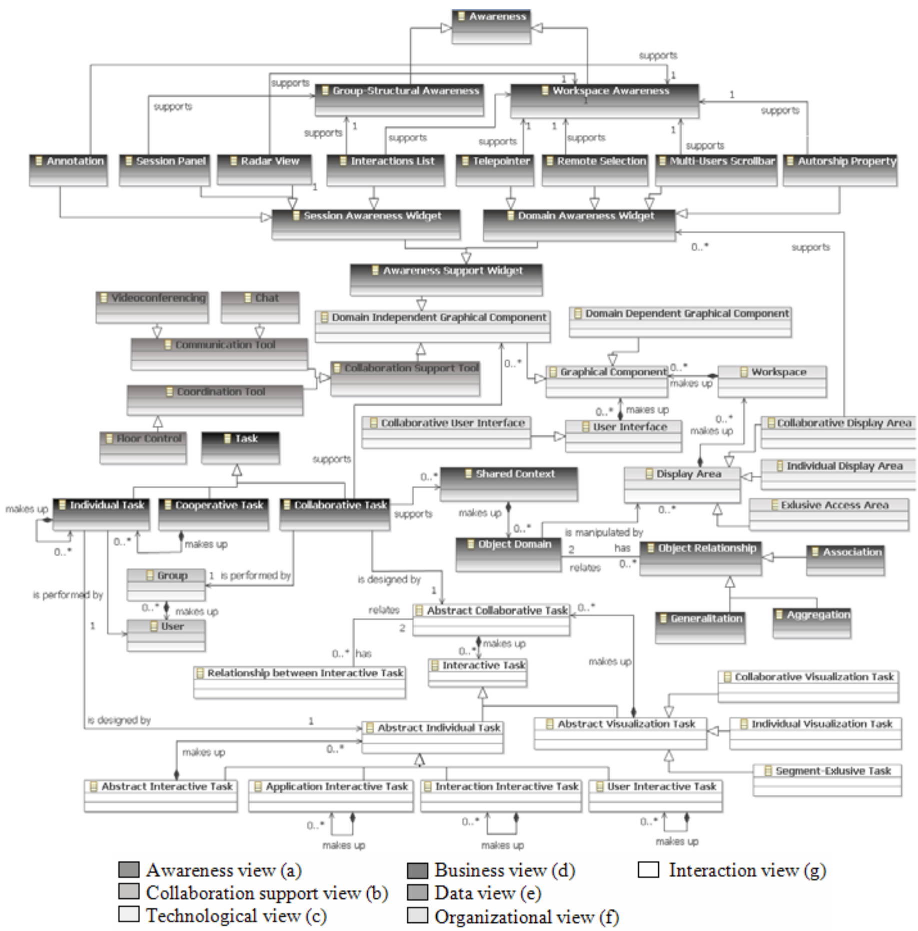


Fig. 1. The Proposed meta-model

Table 1. Definition of Widgets for Awareness Support

Widget	Type	Description
Radar View	*Session Awareness Widget*	Widget that shows the area in which users are working and the activities they are undertaking.
Telepointer	*Domain Awareness Widget*	Graphical cursor that indicates the position of the mouse pointer of a user who is interacting with the shared context.
Authorship Property	*Domain Awareness Widget*	Widget that shows, for each object in the shared context, the user who created it.
Interactions List	*Session Awareness Widget*	Widget that shows, textually, the interactions taking place during a work session.
Multi-User Scroll Bar	*Domain Awareness Widget*	Widget that shows the location of each user in the work area according to the position of the scroll bar.
Annotation	*Session Awareness Widget*	Brief text note which a user can publish in order to share messages asynchronously.
Session Panel	*Session Awareness Widget*	Widget that shows the actors who are involved in a session, with additional information about them.
Remote Selection	*Domain Awareness Widget*	Widget that shows objects in the shared context that have been selected by other users.

The *Organizational view* (Figure 1.f) includes concepts related to the group work members; the *Data view* (Figure 1.e) includes concepts related to the domain and to the data manipulated within the shared context; the *Technological view* (Figure 1.c) considers the implementation of awareness mechanisms and interactive activity issues, such as the components and workspaces which may go into making up the application's GUI; the *Collaboration support view* (Figure 1.b) includes the concepts related to communication and coordination support; the *Interaction view* (Figure 1.g) includes the concepts related to the modeling of interactive issues in the application; the *Business view* (Figure 1.d) includes the entities involved in the design of tasks, regardless of the technology used; and, finally, the *Awareness view* (Figure 1.a) includes those concepts which are directly related to awareness support.

Awareness support at the implementation level results in the inclusion of graphical elements (widgets) in the user interface. Table 1 shows the widgets for awareness support considered in our proposal, along with a brief description of each of them. We have classified the awareness support widgets into two different types: Session Awareness Widgets and Domain Awareness Widgets. These widgets and the classification, which considers what kind of information is handled in each case, are part of the *Awareness view* in the meta-model (Figure 1.a).

3 Conclusions and Future Work

In this paper we have introduced a meta-model which allows for the identification of those concepts related to awareness which should be considered in any proposal for the modeling of interactive groupware systems, such as the proposal we ourselves have worked with: that of the CIAM approach. Once these concepts had been identified, along with the relationships between them, the next step was to define a graphical notation for modeling them. In our case, we opted to take the CIAN graphical notation as our starting point.

Our work in the future will be to define the visual appearance of the notation and to implement a software tool with support for modeling related to our meta-model. This implementation will be developed using meta-model-based graphical editor technologies. Specifically, these will be technologies such as EMF[1] (Eclipse Modelling Framework) and GMF[2] (Graphical Modelling Framework). By using these technologies, it will be possible to test the semantics of the generated models, and to validate the proposed meta-model. This tool will support the semi-automatic generation of collaborative interfaces from CIAN models and interaction models, both enriched with awareness support elements.

Acknowledgements. This work has been partially supported by the iColab project, funded by the Junta de Comunidades de Castilla-La Mancha (Spain), and by the Prog-CoLab project, funded by the Universidad de Castilla-La Mancha (Spain).

References

1. Dourish, P., Bellotti, V.: Awareness and coordination in shared workspaces. In: Proceedings of the 1992 ACM Conference on Computer-Supported Cooperative Work, Canada (1992)
2. Gutwin, C., Greenberg, S.: A Descriptive Framework of Workspace Awareness for Real-Time Groupware. Computer Supported Cooperative Work 11(3), 411–446 (2002)
3. Gallardo, J., Molina, A.I., Bravo, C., Redondo, M.A., Collazos, C.: An ontological conceptualization approach for awareness in domain-independent collaborative modeling systems: Application to a model-driven development method. Expert Systems with Applications 38(2), 1099–1118 (2011)
4. Kokar, M.M., Matheus, C.J., Baclawski, K.: Ontology-based situation awareness. Information Fusion 10(1), 83–98 (2009)
5. Molina, A.I., Redondo, M.A., Ortega, M.: A Review of Notations for Conceptual Modeling of Groupware Systems. In: New Trends on Human-Computer Interaction, pp. 1–12. Springer, London (2009)
6. Molina, A.I., Redondo, M.A., Ortega, M.: A methodological approach for user interface development of collaborative applications: A case study. Science of Computer Programming 74(9), 754–776 (2009)
7. Molina, A.I., Redondo, M.A., Ortega, M.: A Conceptual and Methodological Framework for Modeling Interactive Groupware Applications. In: Dimitriadis, Y.A., Zigurs, I., Gómez-Sánchez, E. (eds.) CRIWG 2006. LNCS, vol. 4154, pp. 413–420. Springer, Heidelberg (2006)
8. Ellis, C.A., Gibbs, S.J., Rein, G.: Groupware: some issues and experiences. Communication of ACM 34(1) (1991)

A Longitudinal Pilot Study to Evaluate Non-visual Icons in a Mobile Exertion Application

Huimin Qian, Ravi Kuber, and Andrew Sears

UMBC, 1000 Hilltop Circle, Baltimore, MD 21250, USA
{huimin1,rkuber,asears}@umbc.edu

Abstract. This paper describes an evaluation of a mobile exertion application, developed to promote physical activity among adult users. To extend previous work, we aimed to identify changes in walking behavior over time, and to determine the efficacy of non-visual cues presented via the phone. Data was gathered using three methods: automated logging, participant-maintained diaries and interviews. Findings revealed that participants were able to respond most effectively to multimodal icons to modify their pace over a two week period. However, their preferences were noted to change depending on the situational context. The study has also highlighted the complexities associated with data collection techniques for mobile evaluations.

1 Introduction

Mobile persuasive applications, such as MPTrain [4] and StepMan [1] have been developed to address the challenges faced remaining motivated to keep fit. However, these are often evaluated at one fixed point in time, under controlled conditions (e.g. lab-based). In contrast, longitudinal studies can provide a more comprehensive overview of usage and an insight into behavioral patterns. Examples include the three-week evaluation of Ubifit Garden [3], where the researchers were able to monitor the type and duration of activities conducted, and how these varied over time. Reis et al. [6] were able to identify the contexts in which modalities were selected for use when using mobile technologies. User preference was found to change depending on the environment, and other factors such as lightning, movement, and hand usage.

In this paper, we briefly describe the development of a mobile exertion application designed to aid users to maintain and modify levels of pace, through the presentation of non-visual icons. To extend previous work [5], and as a first step towards understanding user behavior, a pilot study was conducted where participants used the application over a two week period in locations of their choice. The aim was to identify the following:

- Which types of feedback are most commonly used to maintain levels of physical activity when walking in uncontrolled environments?
- Does choice of feedback vary depending on environment and/or context?
- How do participants behave in response to signals rendered via different modalities?

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 458–461, 2011.

2 Design of Mobile Application and Pilot Study

A pedometer-style application has been developed for the Nokia N95 mobile telephone. The device is strapped to the waist using a belt. As the user walks around his/her environment, the number of steps are calculated using data gathered from the built-in accelerometer. An optimal level of pace is determined, based on the user's step rate. Auditory, tactile or multimodal (composed of both auditory and tactile effects) prompts (all played at 158 Hz) are presented via the phone, to assist the user in maintaining his/her desired rate of pace by encouraging him/her to modify pace as appropriate. To move faster, the user is presented with icons lasting for 200ms, while to move slower, 800ms icons are played. The non-visual cues are presented every six seconds until the user reaches their target rate of pace. All auditory information was presented at 65dB. To extend our previous work, a study was performed to determine the efficacy of non-visual cues over a two week period while traversing a variety of terrains, rather than being constrained to a controlled environment, such as in [5]. We also aimed to identify whether over time, users would react faster to modify their level of pace.

3 Methodology

Three adults (2 male, 1 female), aged between 50 and 60, were selected from a local community center. No difficulties were reported with levels of hearing or mobility. All participants classified themselves as moderately physically active. However, the time spent weekly on walking would depend on work schedules and other commitments. Participants were each provided with thirty minutes of training with the application, during which time their pace was recorded when walking at a comfortable speed (N). This enabled us to determine a desired level of pace (N * 120%). Non-visual icons would be presented by the application, if participants were unable to maintain their individual target paces. Participants were then asked to use the mobile application at any point over the two week period, and record their walking goals and experience of using the application in an online diary. The distance and time taken for each journey, rate of pace, and how effectively users were able to maintain this pace was automatically logged by our software. We envisaged that recording data in a diary would aid us in gaining a deeper understanding of their walking behavior, by providing an insight into key events within the journey and reasons behind the selection of feedback. It would also reduce the risk of problems associated with recalling past events, as these would be noted immediately after each trial. Online diaries could also be monitored by the evaluator without disturbing the users.

At the end of the first and second week of use, the participants were interviewed to follow-up on events logged within the diaries and to obtain step rate data stored on each mobile handset. The interviews focused on which types of feedback had been used, key events that disrupted any walk, frequency of use, levels of workload experienced, and if participants had faced any difficulties with the application.

4 Results and Discussion

Ten trials were logged by the three participants, lasting between seven to thirty-five minutes in duration (Table 1). All trials were conducted in outdoor environments, where visual distractions were present (e.g. traffic) on a range of terrains including both flat and uneven ground (e.g. grass, inclines etc).

Table 1. Results from Study

Trial	Duration (Mins)	Cue	Terrain
1	30	Multimodal	Forest trail
2	35	Tactile	Uneven terrain
3	10	Multimodal	Near home environment
4	10	Multimodal	Uneven terrain
5	7	Multimodal	Uneven terrain
6	9	Multimodal	Uneven terrain
7	15	Tactile	Uneven terrain
8	20	Audio	Near home environment
9	20	Multimodal	Near home environment
10	20	Multimodal	Near home environment

Multimodal icons were self-selected for presentation in seven out of the ten trials. When asked to discuss reasons behind the selection of these cues, participants suggested that combining auditory and tactile feedback together was found to offer a more “noticeable” alert in inclement environments. Participant #1 suggested as the sound of traffic became louder, the audio was more difficult to process, but he was still able to perceive the tactile cues presented while keeping his eyes free to monitor the presence of cars in close proximity to his route. When in quiet environments (e.g. back yard), Participant #3 suggested that she was able to use auditory feedback. However, she suggested in unfamiliar surroundings more feedback would be needed, as more concentration would be spent navigating the paths. It appeared that context and awareness of the environment played a role in selection.

In order to identify how effectively participants were able to react to the non-visual feedback presented, the number of instances when icons were played were calculated for each trial. A total of 120 instances of feedback were presented (11 audio icons, 31 tactile icons and 78 multimodal icons). The majority of feedback was presented towards the beginning of each trial (>85%), mostly aiming to prompt participants to walk faster. This was expected, as participants noted that they would try to walk faster than their comfortable pace, but it was difficult to gauge from memory, what this rate would be. To determine the accuracy of responses to feedback presented, the number of correct responses to the cues, was divided by the total number of six second intervals within each journey. For all walking trials, response accuracy rates were higher than 97%, demonstrating that participants were able to react to all forms of feedback effectively. Due to the small sample of participants, it is difficult to draw conclusions as to whether perception improved during the time period. In all ten trials, the participants walked faster than their desired target pace for certain periods of time. When asked for the reasons behind walking faster than necessary, participants suggested that in order to divide attention between the non-visual cues and the surrounding environment, it was thought to be less cognitively intensive to walk faster. Participant #1 mentioned that he tended to walk faster deliberately to avoid the delivery of signals. However, no participants reported experiencing levels of overload from presentation of the icons.

While participants did not make as much use of the application as initially envisaged, results provided us with awareness of the contexts and environments in which different forms of feedback proved to be useful. In a future study, we aim to enhance the application, to modify the type of feedback presented during each trial, to provide more flexibility when environments change (e.g. when moving from quiet to noisy surroundings). We will then use the longitudinal study to evaluate how users respond to the feedback presented over time.

5 Practicalities of Conducting the Pilot Study

While laboratory-based studies offer an effective way to gather data, conclusions derived from studies may only pertain to performing tasks under constrained conditions. Diary studies were selected as one method to gather data when performing mobile tasks, as it was thought that participants could undertake activities in the field, and would be able to describe their goals and actions with minimal intervention from the evaluators. By being able to view online diaries, it was noticeable that few entries were made. It also required effort to understand the meaning associated with the comments made. As a result, the evaluators needed to contact the participants to follow-up on certain points. It also proved challenging to cross-reference entries from the diary with the steps taken, as comments were not as extensive as anticipated. Certain events were omitted as they were thought to be trivial by participants, also mentioned by [2]. Findings from our study suggested that follow-up interviews provided a richer representation of events, than relying on the diary study alone.

References

1. Bieber, G., Diener, H.: StepMan – A New Kind of Music Interaction. In: Human Computer Interaction International (2005)
2. Brandt, J., Weiss, N., Klemmer, S.R.: txt4l8r: Lowering the Burden for Diary Studies under Mobile Conditions. In: CHI EA 2007, pp. 2303–2308. ACM Press, New York (2008)
3. Consolvo, S., McDonald, D.W., Toscos, T., et al.: Activity Sensing in the Wild: A Field Trial of Ubifit Garden. In: CHI 2008, pp. 1797–1806. ACM Press, New York (2008)
4. Oliver, N., Flores-Mangas, F.: MPTrain: A Mobile Music and Physiology Based Personal Trainer. In: MobileHCI 2006. ACM Press, New York (2006)
5. Qian, H., Kuber, R., Sears, A.: Maintaining levels of activity using a haptic personal training application. In: CHI EA 2010, pp. 3217–3222. ACM Press, New York (2010)
6. Reis, T., Sá, M.D., Carriço, L.: Multimodal Interaction: Real Context Studies on Mobile Digital Artefacts. In: Pirhonen, A., Brewster, S. (eds.) HAID 2008. LNCS, vol. 5270, pp. 60–69. Springer, Heidelberg (2008)

A Remote Multi-touch Experience to Support Collaboration between Remote Museum Visitors

Ernesto Arroyo, Valeria Righi, Roger Tarrago, and Josep Blat

GTI, ICT Department, Universitat Pompeu Fabra,
Roc Boronat 138, 08018 Barcelona, Spain
{ernesto.arroyo,valeria.righi,roger.tarrago,josep.blat}@upf.edu

Abstract. This paper presents a collaborative experience designed to support learning in two remotely located museums sharing a common exhibition. A remote collaborative multi-touch experience offers an additional channel for museum visitors to explore the exhibition and increase the sense of connectedness and awareness between the two spaces. The experience flow includes stages offering opportunities for exploration, negotiation and cooperation. The paper describes the design and implementation of a system that allows simultaneous collaborative interaction and communication through two multi-touch surfaces augmented with videoconferencing. The system allows museum visitors to communicate with remote participants and with their peers. Finally, the paper discusses preliminary observations of end-users, and cultural organizations using the prototype. This work provides a use case for social interactive experiences that could draw museum visitors to further explore an exhibition and share their views and interpretation with others.

Keywords: Computer supported collaborative learning, multi-user interaction, informal learning, serious games, remote awareness, interactive surfaces.

1 Introduction

This project is framed within the context of two Spanish museums that exhibit an art collection simultaneously in Figueres and Barcelona cities. Each museum exhibits part of a collection with the intention to allow visitors to freely explore and experience the collection. These conditions offer interaction designers a unique opportunity to explore different ways for inter-connecting museums and its visitors using a common exhibition to build the discourse. This paper presents the design and implementation of a prototype designed to support collaboration and learning in remote museum environments. The prototype facilitates *simultaneous collaborative interaction* and *communication* through two multi-touch surfaces and a videoconferencing system; similar to the DigiTable project [7].

1.1 Related Work

Several museum guides, tools, and games have been developed to support location-aware collaboration, cooperation and interaction using devices that could be located

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 462–465, 2011.

anywhere within a museum [3][4]. However, few have been specifically designed for inter-connecting two remote museum spaces, e.g. [2]. On the other hand, online collaborative systems include videoconferencing and digital sharing capabilities that are already used in educational and cultural settings [10][9]. Additionally, tabletops have also been used to facilitate face-to-face collaborative learning. These tools allow simultaneous collaboration between multiple remote participants, which facilitate the provision of learning [5], support online collaboration activities [6] [7] and improve the cognitive proximity and identity decoupling of participants. Our work incorporates several technologies to make explicit the relationship between two "partner" museums hosting a common exhibition through a system composed of two stations placed in two remote locations.

2 Remote Multi-touch Experience

Together, with museum stakeholders, we designed a system targeted to students groups visiting a museum during a coordinated activity in two distant cities. The system seeks to offer an additional channel for museum visitors to explore the exhibition and increase the sense of connectedness between the two spaces by allowing museum visitors to interact and collaborate. Each station includes two 32' screens: one horizontal multi-touch screen, and a vertical screen (Fig. 1). The horizontal screen incorporates a multi-touch table that allows for direct manipulation of up to 32 objects and displays content related to the exhibition that can be seen and accessed simultaneously from both stations. The vertical screen displays a video feed from each location and complementary information related to the learning experience.

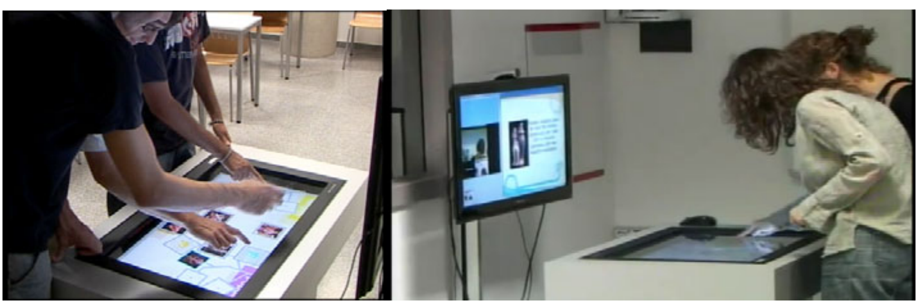


Fig. 1. The system includes two remotely located stations. Participants interact on a horizontal surface while sharing the same workspace with their remote partners. *Left*) co-located participants interacting simultaneously. *Right*) a vertical screen shows a video feed of the remote location facilitating communication between the two locations.

2.1 Experience Description

Groups of participants are invited to play a "game" with other groups also visiting a museum at a different city. The game is related to the objects presented in both museums. During the game, groups of participants interact with each other as each group holds key information related to the exhibition in their own cities that will help the other group complete the game tasks. The experience flow is modeled in three stages that offer different opportunities for exploration, negotiation and cooperation and that guide and facilitate discovering and exploring the installation materials. The first stage provides useful information for completing the other stages and it promotes negotiation by requiring participants to coordinate the material to be explored. The other two stages include focused collaborative tasks [1] that can be solved by collaborating with other participants and encourage information sharing between visitors. The system supports simultaneous manipulation over a shared collaborative workspace for users located remotely and users co-located. Remote users share the same visual perspective using a *what you see is what I see* interaction metaphor [8]. Several visual cues help participants understand the user actions in the remote museum. Finally, the system includes a videoconference service to promote communication and reflection between remote groups [1]. The aim is that museum visitors will use the videoconference to explain, and justify their actions over the virtual shared space.

3 Preliminary Evaluation

The experience went through an iterative design process where high school and university students informally tested several prototype versions. The experience has also been presented to several cultural and governmental organizations in two public events. The game lasted less than 10 minutes on average and we observed a learning curve regarding simultaneous interactions; on the first stage participants exhibited a turn-taking behavior, while on the second stage they interacted with the objects simultaneously and independently. We also observed a similar curve in regard to remote collaboration, which decreased as the game advanced. The videoconferencing proved useful as participants used it in a natural way to communicate with remote participants. Visual cues used to facilitate remote awareness proved useful in supporting simultaneous remote interaction; however, some participants claimed that the workspace seemed chaotic when all participants were moving items. Finally, we also found that individuals wanted to experience the installation by themselves without having a remote partner on the other end. This can be easily addressed by offering single and collaborative modes of interaction in future iterations.

4 Conclusion and Future Work

We have presented an informal learning experience designed to connect two remotely located museums sharing a common installation on. Our system allows museum visitors to communicate between and with their peers around a multi-touch tabletop through videoconferencing. We have validated that the system supports natural interaction and allows multiple visitors to interact simultaneously on a tangible shared space while also facilitating face-to-face collaboration among participants within the same location, and among remote participants. This project illustrates one possible way for inter-connecting museums and its visitors using a common exhibition. Future work includes evaluating several usability and interaction aspects of the system, identifying collaborative tasks suitable for this type of settings, and improving reliability issues required for everyday use.

Acknowledgments. This work has been partially funded by the Learn 3 project, (TIN2008-05163/TSI). The authors would like to thank the cultural and technical partners of the C3I2 workgroup for their support and ideas.

References

1. Baker, M., Lund, K.: Promoting reflective interactions in a CSCL environment. Journal of Computer Assisted Learning 13(3) (2003)
2. Galani, A., Chalmers, M.: Empowering the remote visitor: Supporting social museum experiences among local and remote visitors. In: 2nd International Conference of Museology (ICM), Mytiline, Greece (2004)
3. Ghiani, G., Paternò, F., Santoro, C., Spano, L.D.: UbiCicero: A location-aware, multi-device museum guide. Interacting with Computers 21(4), 288–303 (2009)
4. Laurillau, Y., Paternó, F.: Supporting Museum Co-visits Using Mobile Devices. In: Brewster, S., Dunlop, M.D. (eds.) Mobile HCI 2004. LNCS, vol. 3160, pp. 451–455. Springer, Heidelberg (2004)
5. Looi, C.-K., Chen, W., And Ng, F.-K.: Collaborative activities enabled by GroupScribbles (GS): An exploratory study of learning effectiveness. Computers & Education 54(1), 14–26 (2010)
6. Palloff, R.M., Pratt, K.: Collaborating online. Learning together in community. Jossey-Bas, San Francisco (2005)
7. Pauchet, A., Coldefy, F., Lefebvre, L., et al.: Tabletops: Worthwhile Experience of Collocated and Remote Collaboration. In: Proc. TABLETOP 2007, pp. 27–34 (2007)
8. Stefik, M., Bobrow, D.G., Foster, G., Lanning, S., Tatar, D.: WYSIWIS revised: early experiences with multi-users interfaces. ACM Trans. Inf. Syst. 5(2), 147–167 (1987)
9. Vom Lehn, D., Heath, C., Hindmarsh, J.: Exhibiting Interaction: Conduct and Collaboration in Museums and Galleries. Symbolic Interaction Journal 24(2), 189–216 (2001)
10. Wunar, B., Kowrach, N.: Live...from the Heart: Using Interactive Videoconferencing to Link Classrooms to Museum Resources. In: Gibson, I., et al. (eds.) Society for IT & Teacher Education International Conference, pp. 581–582 (2009)

A Systematic Evaluation of Mobile Applications for Diabetes Management

C. Martin, D. Flood, D. Sutton, A. Aldea, R. Harrison, and M. Waite

Oxford Brookes University, UK
{cemartin,derek.flood,dsutton,aaldea,rachel.harrison,
mwaite}@brookes.ac.uk

Abstract. This short paper contains a summary of work that is currently in progress towards the development of an intelligent, personalised tool for diabetes management. A preliminary part of the development process has consisted of a systematic evaluation of existing applications for mobile phones.

Keywords: Efficiency, effectiveness, keystroke level modelling, heuristics.

1 Introduction

The combined technologies of mobile phones and cloud computing have huge potential to improve health care services [1], but a crucial factor in the adoption of such services is the usability of the software [2, 6]. Previous studies have been conducted to evaluate the effectiveness of mobile phones for assisting with diabetes self-management, eg [12], and standards are beginning to emerge with regard to the design of mobile applications in general [14] as well as mobile medical applications in particular (eg ISO/IEC 62366), but research into the usability evaluation of applications for mobile devices is still a relatively new area [7]. Some issues that affect the usability of mobile applications include mobile context, connectivity, screen size and resolution, limited processing power and data entry methods [15], and these issues remain broadly the same despite recent advances in mobile phone technology. The purpose of this study is to uncover usability issues in order to know what pitfalls to avoid in the design of an application for diabetes management, and also to determine what functionality is typically included in such applications.

Type 1 diabetes (T1DM) occurs when the insulin producing cells of the pancreas are destroyed leaving the body unable to control its blood glucose levels. People with T1DM have to take insulin regularly to try to keep their glucose levels within a safe range. The vast majority of patients with T1DM in the UK administer their insulin through multiple daily injections, and the remaining proportion use insulin pumps. Modern insulin pumps include dose calculators to help patients determine how much insulin to administer, but people on multiple daily injections do not usually have this support, and tend to do the calculations themselves. The resulting need for electronic decision support, combined with the recent growth in smart phone use has led to the development of a plethora of diabetes management applications, despite the many barriers that have inhibited the adoption of electronic systems for diabetes management in the past [3,8]. One of the key barriers has been usability, and another

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 466–469, 2011.

has been the economic implications for healthcare providers, but the standardisation of controls and rigid human interface guidelines imposed by the providers of most phone platforms has already had a huge impact on usability, and the ease with which software applications for mobile phones can now be developed and distributed helps to mitigate against the economic factors. The usability issues listed in [15], are all applicable to systems for diabetes management, but in addition, common complications of diabetes, such as retinopathy and peripheral neuropathy, could affect a patient's ability to view data or use a touch screen respectively.

2 The Evaluation

The protocol used here to analyse the applications is fairly general and has also been applied to mobile spreadsheets [9]. The results given below are limited to the iOS platform, but work is in progress to extend the evaluation to other platforms. The number of spreadsheet applications available for iOS is only a fraction of those for diabetes management, but some similar issues emerged with data entry and visualisation for example. The following steps were performed:

1. *Identify all potentially relevant applications:* this involved a search on the iPhone app store using the keyword *diabetes* which returned a total of 231 applications.
2. *Remove light or old versions of each application.* There were 9 light versions of applications which were removed from the list, leaving 222 apps.
3. *Identify the primary operating functions and exclude all applications that do not offer this functionality.* The key functionality was mapped to corresponding tasks which are shown in Table 1. Only 8 applications had all of the functionality.

Table 1. Tasks to evaluate primary functionality

Task	Description	Task	Description
1	Set measurement units	4	Log insulin dose
2	Log blood glucose level	5	Display data graphically
3	Log carbohydrate intake	6	Export data via email or similar

4. *Identify all secondary functionality*. A full list of additional functionality that was offered by the various applications was compiled to be assessed by patients and clinicians in the second phase of this study.

5. *Construct tasks to test the key functionality using each of the methods below:*
   a. Keystroke level modelling [4,13] was performed on the final 8 applications using the tasks from Table 1. The number of times the screen was touched in order to complete each task was recorded, where a touch might refer to a single touch or a swipe of the finger. The results are shown in Table 2.

Table 2. Results of KLM

Application Task:	1	2	3	4	5	6	Total
RapidCalc Insulin Dose Manager	10	2	1	3	2	1	19
GluCoMo	5	5	4	4	1	3	22
Diabetes Diary	5	3	3	2	3	6	22
Diabetes Personal Manager	3	7	4	3	3	6	26
DiabetesPlus	2	7	6	6	1	6	28
LogFrog DB	6	5	6	5	1	7	30
Diabetes Buddy - Control your Blood Sugar	3	9	10	7	1	8	38
Diabetes Pilot	5	5	7	3	3	27	50

This analysis excluded the actual time taken to complete tasks, which will be measured in the study on patients. The KLM process highlighted a number of key differences between the applications that could have a significant effect on usability. For example, the data logging occurs so frequently that it is essential for it to be as efficient as possible, and yet a number of different methods were in use including sliders, pickers and various styles of keyboard. Some applications also made much more effective use of default settings than others.

b. Heuristics [10] were then used to identify more usability problems. The four applications that scored the best in the KLM were subjected to heuristic evaluation by five expert evaluators, using guidelines that take into account specific issues associated with mobile systems [3]. Each heuristic was ranked using Nielsen's five-point Severity Ranking Scale (SRS) [11].

Table 5. Average heuristic ranking

Heuristic:	A	B	C	D	E	F	G	H	Tot
RapidCalc	1.1	0.9	0.4	0.4	0.8	1.3	1.3	2.0	8.2
GluCoMo	0.6	1.4	1.2	1.7	0.8	1.3	1.3	1.7	10
Diabetes Diary	0.5	0.8	0.9	0.5	0.9	0.7	1.4	0.8	6.5
Diabetes Personal Manager	0.6	0.8	0.9	1.1	1.1	0.9	1.6	2.3	9.3

Each evaluator provided evidence for each ranking, and this exposed some problems that had not been shown by the KLM. For example, there were problems with visibility of battery status, and limited ability to restore data if the device was lost. There were some navigation problems too, but some of these were only learnability issues. Some apps were over-crowded but most offered some degree of personalisation. There were some serious problems with error management and security: it was not easy to edit records on most apps, and data transfer was typically done via unencrypted email, which is not appropriate for healthcare information. Unfortunately it is beyond the scope of this paper to give a thorough analysis of the heuristic evaluation, but that will be the subject of a future paper.

3 Conclusion and Future Work

This short paper has described a systematic evaluation of diabetes management applications for the iOS platform. The KLM evaluation highlighted a number of usability issues associated with data entry and personal settings, and the heuristic evaluation exposed further problems associated with loss of the device, learnability, aesthetics, error management and security. Future work includes a usability study on patients using additional mobile platforms, and the ultimate goal is to design an application for diabetes management with a decision support engine and secure data transfer taking into account the associated legal implications. Further work is also planned on keystroke level modelling of mobile applications, in order to standardise measurements such as what constitutes a single touch or swipe.

References

1. Alagöz, F., Calero Valdez, A., Wilkowska, W., Ziefle, M., Dorner, S., Holzinger, A.: From Cloud Computing to Mobile Internet, From User Focus to Culture and Hedonism: The Crucible of Mobile Health Care and Wellness Applications. In: IEEE 5th International Conference on Pervasive Computing and Applications, pp. 38–45 (2010)
2. Bellazzi, R.: Telemedicine and Diabetes Management: Current Challenges and Future Resarch Directions. Journal of Diabetes Science and Technology 2(1), 98–104 (2008)
3. Bertini, E., Catarci, T., Dix, A., Gabrielli, S., Kimani, S., Santucci, G.: Appropriating Heuristic Evaluation Methods for Mobile. International Journal of Mobile Human Computer Interaction 1(1), 20–41 (2009)
4. Card, S.K., Thomas, T.P., Newall, A.: The Psychology of Human-Computer Interaction. Lawrence Erbaum Associates, London (1983)
5. Calero-Valdez, A., Ziefle, M., Firat, A., Holzinger, A.: Mental Models of Menu Structures in Diabetes Assistants. In: Miesenberger, K., Klaus, J., Zagler, W., Karshmer, A. (eds.) ICCHP 2010. LNCS, vol. 6180, pp. 584–591. Springer, Heidelberg (2010)
6. Greenhalgh, T., Hinder, S., et al.: Adoption, non-adoption, and abandonment of a personal electronic health record: case study of HealthSpace. BMJ 341, c5814 (2010)
7. Kjeldskov, J., Stage, J.: New techniques for usability evaluation of mobile systems. International Journal of Human Computer Studies 60(5-6), 599–620 (2004)
8. Klonoff, D.: Using Telemedicine to Improve Outcomes in Diabetes - An Emerging Technology. Journal of Diabetes Science and Technology 3(4), 624–628 (2009)
9. Martin, C., Flood, D., Harrison, R.: A Protocol for Evaluating Mobile Applications. To appear in IADIS (2011)
10. Nielsen, J., Molich, R.: Heuristic evaluation of user interfaces. In: The Conference on Human Factors in Computing Systems: CHI 1990 (1990)
11. Nielsen, J.: Usability Engineering. Morgan Kaufmann Publishers, San Francisco (1994)
12. Preuveneers, D., Berbers, Y.: Mobile phones assisting with health self-care: a diabetes case study. In: Proceedings MobileHCI 2008 Proceedings of the 10th International Conference on Human Computer Interaction with Mobile Devices and Services (2008)
13. Schulz, T.: Using the Keystroke-Level Model to Evaluate Mobile Phones. IRIS 31 (2008)
14. W3C. Mobile Web Application Best Practices, `http://www.w3.org/TR/mwabp/`
15. Zhang, D., Adipat, B.: Challenges, Methodologies, and Issues in the Usability Testing of Mobile Applications. International Journal of Human-Computer Interaction 18(3), 293–308 (2005)

An Integrated Approach to Develop Interactive Software

Begoña Losada, Maite Urretavizcaya, and Isabel Fernández de Castro

Dept. of Computer Languages and Systems,
Faculty of Computer Engineering,
University of the Basque Country,
20001 San Sebastián
{b.losada,maite.urretavizcaya,isabel.fernandez}@ehu.es

Abstract. In this poster we present InterMod, an approach that combines Agile Methods, Model-Driven Developments and User-Centered Design, which are widely accepted in the development of interactive software. The planning and project organizing are based on User Objectives (user desires). The project is organised as a series of iterations and the work is distributed in different workgroups according to some developmental and integration activities. The requirements are incrementally collected and validated with models based on user-centered design. To speed up this validation, we put forward the SE-HCI model, which enriches a human-computer interaction model with the semantics of the application and some basic characteristics of an abstract prototype.

Keywords: User-Centered Design, Agile methods, Model-Driven Development, Software Engineering.

1 Introduction

Agile methods (AM), model-driven developments (MDD) and user-centred design (UCD) are three approaches widely accepted by the community and share a common objective of efficiency in the resulting software. However, none of them alone achieves success without encountering problems during application development. Because of that, efforts are being made to integrate these techniques so that the advantages of one mitigate the problems of the others [1][2][3][4]. However, due to the fact that a majority of software engineering development processes focus on software architecture, satisfactory integration has not yet been achieved. Therefore, we focus our efforts on a new approach to improve software development that combines agile characteristics, MDD and user-centred techniques.

2 InterMod, an Integrated Proposal

InterMod [5] is a methodology whose aim is to help with the accurate development of interactive software. Although it is suitable for use with web design, its utility is not restricted to just that area.

Our proposal, shown in Fig 1, is to organise the project as a series of iterations, just as the agile methodologies do, and distribute the work in the iterations according to

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 470–474, 2011.

different developmental activities of the *User Objectives* (UO). A *User Objective* is a user desire e.g. "*buying a t-shirt*" or "*reserving a meeting room in a workplace*", that can be achieved by one or more user functionalities.

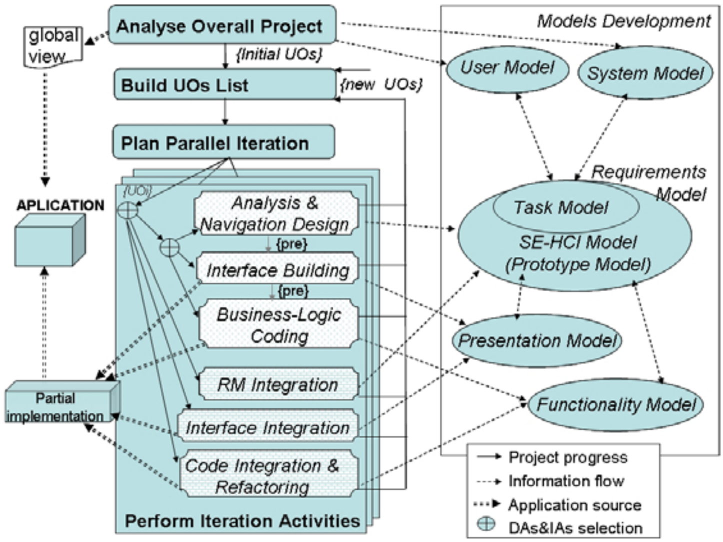


Fig. 1. InterMod process and development activities

The new UOs are in turn objectives to be refined in subsequent iterations. Project progress is reflected in the activities done to achieve these UOs current and the resulting models.

2.1 InterMod Activities and Models

InterMod has two kinds of ***Activities***: *Developmental Activities* and *Integration Activities*. We represent graphically Activities as shown in Table 1.

Table 1. InterMod Activities

Development Activities	Integration Activities
A1. Analysis & Navigation Design	**I1**. RM Integration
A2. Interface Building	**I2**. Interface Integration
A3. Bus.-Logic Coding	**I3.** Code Integration & Refactoring

The ***Developmental Activities*** (DAs) associated with each UO are strongly related: A1.*Analysis and Navigation Design,* A2.*Interface Building* and A3.*Business-Logic Coding*. Each UO requires the three DAs to be developed but a prerequisite relation must be done A1<A2<A3 ('<' means prerequisite). Just as UCD recommends, before coding a relevant UO, its interface must be validated. However, unlike UCD, it is not required that the complete application interface be developed before moving to the implementation of the business logic; instead this approach stays framed in the development of one or several UO groups. Furthermore, to assure a correct incremental progress of the project, some ***Integration Activities*** (IAs) are needed: *I1.Requirement Models (RM) Integration*, *I2.Interface Integration* and *I3.Code Integration & Refactoring*. A restriction is necessary for controlling the correct development of an IA. Thus, it is possible to carry out an IA I_k (K=1..3) for a concrete UO_j (j=0..n) *if and only if* the UO_j is the fusion of two UOs belonging to the UO List *and* the DAs A_k of these fused UOs are already made. To ensure consistency in the final application, evaluations of the incrementally obtained products as well as heuristic and metric evaluations are included in all activities.

All iterations are guided by the same action plan that divides the work according to the activities of different UOs, in such a way that each DA will be next driven by models and all the integration processes can lead to the revision and modification of these models. Even during final integration of the software there may be revisions of all models and new UO can be created. The activities of **analysis and navigation design** and **RM integration** deals with the *Requirements Model* (RM), which includes the *Semantically Enriched Human-Computer Interaction* (SE-HCI) *model*. The *SE-HCI Model,* that incorporates information from the *User* and *System Models,* is an abstract description constructed over the *Task Model*. It also incorporates three essential aspects: a) The description of both the actions that users and the system can carry out at the user interface level during an interactive session [7], and their possible temporal relations, b) The descriptions of the correct and incorrect interactions that represents the semantics of the application and c) The basic visual characteristics, such as colours, sections, button types, etc. We propose the evaluation of the requirements involved in the SE-HCI with an abstract prototype created automatically by transforming the *SE-HCI model*. From this point, the evaluation can be carried out jointly by the designers, customers and developers.

In the **Interface Building activity**, the *Presentation Model* is created for a UO previously designed and evaluated, and the **Interface Integration activity** fuse together the *Presentation Model* of some UOs. The *Presentation Model* of a specific UO settles the graphical elements and others characteristics gathered from the *Requirements Models*. There are several languages for modeling user interfaces widely used and tested, as XIML[7] or UIML [9], and they may be used to reflect this model. Finally, the **Business–Logic Coding** and **code Integration & Refactoring activities** deal with the *Functionality Model* that guides the implementation in a particular programming language. This model inherits the behaviour characteristics from the UO *Requirements Models* evaluated in the first activity. UML or SysML [10] are alternative languages typically used to represent this model.

2.2 InterMod Process Steps

InterMod has four main steps (see Fig.1), i.e. the initial *Analyse Overall* step, and then an iterative process with three steps follow: *Build User Objectives List*, *Plan Parallel Iteration* and *Perform Iteration Activities*.

At the beginning of the project, InterMod proposes the **Analyse Overall Project** step in order to determine: (a) what the starting UOs are, such as those most important or needed, that provide the initial global view of the application, and (b) the *System*

Model and *User Model* that help to collect the defining characteristics of the system type (e.g. device type, security, window size, colour, logo, etc) and those of the user (e.g. colour preferences, font, size, some limitations as colour blindness, deafness, vision loss, etc). All developments in the project will inherit or extend these models in order to guide and to ensure coherence throughout the entire application.

The application requirements are incrementally collected during the progressive UO List construction. Each iteration begins with a revision of the UOs list. The ***Build User Objective List*** step updates the list with the new UOs derived either from the previous UOs developments or from the new needs of the project. That is, the UOs included in the list may be modified, in the sense of agile methodologies [5], through the different evaluations undertaken by developers and users, or by the continuous meeting among members of the same and different teams. In order to achieve an UO, different activities must be realized. The next step, ***Plan Parallel Iteration***, decides for the current iteration: (a) what UOs to develop, (b) what activities to make for those UOs and (c) how to distribute these different activities to the workgroups (if there is more than one). The iteration ends with the ***Perform Iteration Activities*** step. Each workgroup performs the activities established in its plan.

A snapshot of a **Project Progress State** and the Plan obtained for the **Parallel Iteration** after some iterations (iteration i) are shown in **Fig 2.** Three aspects characterize the state of the project: the UO list, the UOs fusion list and the UOs progress according to theirs ***Activities*** made. In this iteration, the ***Parallel Iteration Plan*** has been the follow: The first team takes responsibility for two activities: **A1** activity for UO_6 and **I1** for UO_4. As it is shown, UO_4 is the fusion of the OU 2 and 3.The team 2 must build the interface (A2) for the UO_1 whose prerequisite is reached (UO_1 is in the A1 list). Meanwhile, team 3 must integrate and refactor the code referred to UO_{10} that is composed of the OU 0 and 5 that have been already coded.

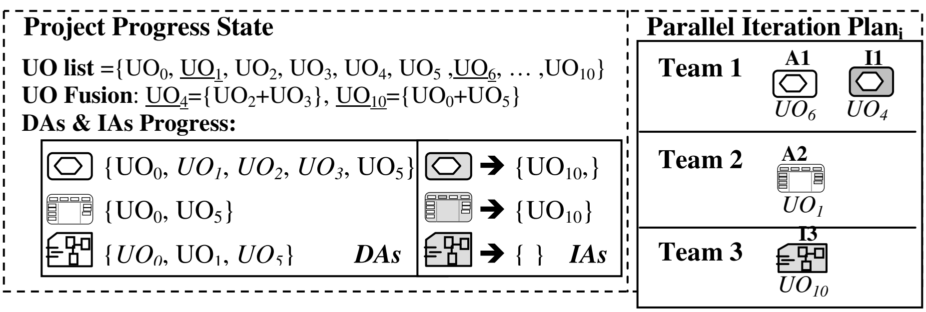


Fig. 2. A Snapshot of the Project Progress with InterMod

3 Conclusions and Future Work

In this poster we present a new vision of the InterMod methodology, a proposal integrating three philosophies: UCD, MDD and AM. From the point of view of agile methods, our work is organized in a series of iterations in which the user objectives to be dealt with are developed. This iterative process speeds up the development and gets results of the project progress. InterMod proposes some developmental and integration activities driven by models to achieve the UOs. The possibility to distribute the work in parallel increases the speed of resolution, although the process itself requires integration points to ensure consistency.

This process allows gather and validate the requirements incrementally. Because of this agile approach, InterMod, unlike UCD, does not require the complete development of the application interface before the implementation of the business logic, but assures usability. The SE-HCI model is the core of our proposal models architecture. It is involved in a Model Driven Process that obtains an abstract prototype created automatically by transforming the *SE-HCI model*. This prototype allows the evaluation of the requirements and facilitates the end user's participation, as recommended by UCD and AM. Early evaluations of the requirements reduce the number of the corrections further on in the process and therefore, reduce its cost.

The new InterMod methodology has been refined in parallel with the development of a demonstrator. A small initial set of UOs has evolved to a complex system. This make us think of the scalability and practicability properties of the proposed methodology. However these aspects have not been treated in this paper as a deeper work needs to be done.

Acknowledgements. This work is supported by TIN2009-14380 and DFG 157/2009.

References

1. Robles, E., Grigera, J., Rossi, G.: Bridging Test and Model-Driven Approaches in Web Engineering. In: Gaedke, M., Grossniklaus, M., Díaz, O. (eds.) ICWE 2009. LNCS, vol. 5648, pp. 136–150. Springer, Heidelberg (2009)
2. Ambler, S.W.: The object primer: agile modeling-driven development with UML 2.0. Cambridge University Press, Cambridge (2004)
3. Ferreira, J.: Interaction Design and Agile Development: A Real- World. Pers., Ph.D (2007)
4. Propp, S., Buchholz, G., Forbrig, P.: Integration of Usability Evaluation and Model-based Software Development. Journal Advances in Engineering Software 40(12), 1223–1230 (2009)
5. Losada, B., Urretavizcaya, M., Fernández-Castro, I.: The InterMod Methodology: An Interface Engineering Process linked with Software Engineering Stages. In: New Trends on HCI: Reseach, Develop., New Tools and Methods, Springer, Heidelberg (2009)
6. Larman, C.: Agile & Iterative development: A manager's guide. Addison-Wesley, Reading (2004)
7. Paternò, F.: Model-Based Design and Evaluation of Interactive Applications. Springer, London (1999)
8. eXtensible Interface Markup Language, http://www.ximl.org/
9. Abrams, M., Helms, J.: UIML Specification (2002)
10. Nolan, B., Brown, B., Balmelli, L., Bohn, T., Wahli, U.: Model Driven Systems Development with Rational Products. ibm.com/redbooks (2007)

Analyzing the Level of Presence While Navigating in a Virtual Environment during an fMRI Scan

Miriam Clemente[1,*], Alejandro Rodríguez[1], Beatriz Rey[1,2], Aina Rodríguez[3], Rosa M. Baños[2,4], Cristina Botella[2,3], Mariano Alcañiz[1,2], and César Ávila[3]

[1] Universidad Politécnica de Valencia, I3BH /LabHuman, Camino de Vera s/n, 46022, Valencia (Spain)
[2] Ciber Fisiopatología Obesidad y Nutrición (CB06/03) Instituto Salud Carlos III, Spain
[3] Departamento de Psicología Básica, Clínica y Psicobiología.- Universidad Jaume I de Castellón (España)
[4] Departamento de Personalidad, Evaluación y Tratamientos Psicológicos- Universidad de Valencia (España)
{mclemente,arodriguez,brey,malcaniz}@labhuman.i3bh.es

Abstract. We have conducted an fMRI research using virtual reality to study the level of presence that subjects experience during the navigation through a virtual environment, in comparison with the presence felt during a video or a photograph viewing task. The fMRI results have not been analyzed yet, but responses to presence questionnaires have been analyzed. Presence levels are similar to those obtained while monitoring with other brain imaging techniques. The highest values are obtained for navigation tasks followed by video and photographs tasks.

Keywords: fMRI, presence, virtual reality, navigation, SUS questionnaire.

1 Introduction

Virtual environments can evoke in the subject the feeling of "being there", commonly known as presence, despite your body is physically located in another place [1, 2]. The traditional way of measuring presence has been using questionnaires [2, 3]. Objective techniques have also been proposed, mainly based on psychophysiological measurements, such as skin conductivity and heart rate [4]. Neurological measures such as EEG or Transcranial Doppler (TCD) have also been used for this purpose. Regarding TCD, two recent studies [5, 6] found that the major cerebral arteries showed changes in blood flow velocity associated with different levels of presence in different immersive and navigation conditions.

Functional magnetic resonance (fMRI) has been used combined with VR for different purposes. Pine et al. [7] analyzed the brain areas activated during the performance of a navigation task without explicitly measuring presence. Hoffman et al. [8] measured by questionnaires the level of presence experimented during an fMRI scan, observing that, despite the constraints of the fMRI machine, the illusion of

[*] Corresponding author.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 475–478, 2011.

presence was possible. Baumgartner et al. [1] used fMRI to analyze brain activation associated to presence during a video of a virtual experience.

Our current research is trying to find the brain areas associated to the sense of presence during a virtual reality paradigm where the participants can navigate freely, in comparison with less immersive conditions, using fMRI. The main hypothesis of our research is that brain activation would be higher in a navigation task than in a video or photographs one. Although the fMRI data have not been analyzed yet, we will present the level of presence evaluated by means of a validated questionnaire [3]. These questionnaire results will be compared with those obtained in a previous work developed with TCD using the same SUS questionnaire [5].

2 Methods

This study has been conducted among 14 right-handed women, between 19 and 25 years old (mean 21.643 ± 2.098), all without any medical or psychological disorder. Hand dominance was tested by the Edinburgh Handedness Inventory [9].

The virtual world has been developed using GameStudio software. It consisted in a common bedroom. Each of the experimental conditions is six times repeated in a counterbalanced order, to avoid effects produced by the order of the tasks. A searching task (key counting) has been introduced to avoid that the subjects remain still during the period. The total time of the whole experiment is 12 minutes 52 seconds. Tasks were trained before the experiment. As can be observed in Figure 1, each repetition of the experimental condition (20 seconds) is preceded by a label indicating the condition and followed by a question about the key counting task.

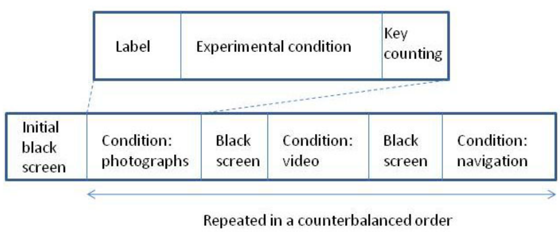


Fig. 1. Diagram of the experimental task

All subjects were scanned during exposure to the different experimental conditions in a 1.5 Teslas Siemens Avanto Magnetic Resonance scanning device (General Hospital, Castellón, Spain). For showing the environments, we used special MRI glasses, VisualStim Digital; and, for the navigation, we used an adapted joystick. First of all, sagittal T1-weighted structural images were acquired, and then the functional scanning was launched synchronized with the virtual environments. Functional images were obtained using a single-shot echo-planar imaging (EPI) sequence.

After the scan, subjects are required to answer six 7-point Likert type questions of a SUS questionnaire [3] to evaluate the level of presence that they have felt during each of the three tasks. The SUS questionnaires [3] were analyzed using SPSS 17.0. We calculated an additional measurement: SUS mean, which is the mean score across the six questions. We applied the nonparametric Friedman Test to compare between SUS responses (dependent variables) in the different experimental conditions (independent variable). Post-hoc tests were made with a Wilcoxon Signed-Rank test with Bonferroni correction.

The results have been compared with the SUS questionnaires responses from 9 women obtained in a similar TCD study [5] that also evaluated differences between two conditions: free and automatic navigation. In that case, the environments were visualized in a CAVE-like system. We applied a repeated measures ANOVA to evaluate the influence on the dependent variable (SUS mean) of the within-subjects factor (navigation versus video) and the between-subjects factor (fMRI versus TCD). The homocedasticy was evaluated with the Levene statistic.

3 Results

We applied the non-parametric Friedman Test to the questionnaire answers, obtaining significant differences between the three experimental conditions in all the questions except question 5. The summary of the questionnaire results for each task and the Friedman results for each question can be seen in Table 1. We conducted *Post-hoc* analyses based on Wilcoxon Signed-Rank Tests over the SUS mean results with Bonferroni correction ($p<0.016$). We found no significant differences between the photographs and the video tasks (Z=-1.174, $p=0.241$). However, there was a statistically significant increment in SUS mean in the navigation vs. photographs trial (Z=-2.805, $p=0.005$) and the navigation vs. video trial (Z=-2.550, P=0.011).

Regarding the evaluation of the influence on SUS mean of the kind of technique used for monitoring brain activation (TCD versus fMRI), we found no significant effect for this factor ($F(1,21)=2.701$, p=0.115). A significant effect was found for the navigation factor ($F(1,21)=11.598$, $p=0.003<0.005$) and no effect was found for the interaction between navigation and monitoring technique ($F(1,21)=0.751$, p=0.396). A power analysis using the G*power3 program [10] showed that a total sample of 42 subjects would have been required to obtain the recommended 80% power in a t test comparison between fMRI and TCD, with alpha set at 0.05 and Cohen's d at 0.8 (large effect size).

Table 2. SUS responses (mean value and standard error of the mean) to the questionnaires and results of the Friedman Test for each question and the mean

Question	Photographs	Video	Navigation	χ^2	p
Question1	3.1429±0.3902	3.7857±0.4470	4.4286±0.4286	16.000	<0.001
Question2	2.7857±0.4824	3.1429±0.4788	3.5000±0.5000	6.750	0.034
Question3	2.0000±0.3145	2.5000±0.4160	3.1429±0.4901	10.903	0.004
Question4	3.1429±0.3759	3.1429±0.4901	4.0714±0.4505	6.450	0.004
Question5	3.4286±0.4412	3.5000±0.4160	4.0000±0.4570	5.250	0.072
Question6	2.7143±0.3696	3.0000±0.5027	3.5000±0.5320	6.067	0.048
SUSmean	2.8693±0.3303	3.1788±0.4037	3.7733±0.4256	12.293	0.002

4 Conclusions

We have analyzed the presence questionnaires results in different experimental conditions (navigation, video and photographs) during a virtual reality scan. We have observed that the maximum experience of presence occurred during the navigation task, with lower rating for the video and minimum for the photographs. Comparing our results with those obtained in a previous research about presence using TCD, we found similar presence levels in both studies, presenting a similar trend between both tasks. In fact, we found significant differences between the conditions for both groups, but no differences between the groups were observed. There is only a trend (that does not reach significance) to higher presence ratings in the TCD research, probably due to the more immersive environment (the CAVE-like configuration) and the less intrusive machine (the TCD probes). The magnetic resonance is noisy, requires you to be still and laid and makes difficult the feeling of "being there".

Our research is the first designed using fMRI to analyze the differences in the sense of presence between navigation, video and photographs. We expect that the fMRI analysis will provide information about the brain areas that are activated during the presence experience associated to a free navigation in a virtual environment.

References

1. Baumgartner, T., Speck, D., Wettstein, D., Masnari, O., Beeli, G., Jäncke, L.: Feeling present in arousing virtual reality worlds: prefrontal brain regions differentially orchestrate presence experience in adults and children. Frontiers in Human Neuroscience 2, 1–12 (2008)
2. Baños, R.M., Botella, C., García-Palacios, A., Villa, H., Perpiñá, C., Alcañiz, M.: Presence and reality judgment in virtual environments: A unitary construct? CyberPsychology & Behavior 3, 327–335 (2000)
3. Usoh, M., Catena, E., Arman, S., Slater, M.: Using presence questionnaires in reality. Presence: Teleoperators and Virtual Environments 9, 497–503 (2000)
4. Meehan, M., Insko, B., Whitton, M., Brooks Jr., F.P.: Physiological measures of presence in virtual environments. Paper presented at: 4th International Workshop on Presence (Philadelphia, Pennsylvania, USA) (2001)
5. Alcañiz, M., Rey, B., Tembl, J., Parkhutik, V.: A neuroscience approach to virtual reality experience using transcranial Doppler monitoring. Presence 18, 97–111 (2009)
6. Rey, B., Alcañiz, M., Tembl, J., Parkhutik, V.: Brain activity and presence: a preliminary study in different immersive conditions using transcranial Doppler monitoring. Virtual Reality 14, 55–65 (2010)
7. Pine, D.S., Grun, J., Maguire, E.A., Burgess, N., Zarahn, E., Koda, V., Fyer, A., Szeszko, P.R., Bilder, R.M.: Neurodevelopmental aspects of spatial navigation: a virtual reality fMRI study. Neuroimage 15, 396–406 (2002)
8. Hoffman, H.G., Richards, T., Coda, B., Richards, A., Sharar, S.R.: The illusion of presence in immersive virtual reality during an fMRI brain scan. CyberPsychology & Behavior 6, 127–131 (2003)
9. Oldfield, R.C.: The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9, 97–113 (1971)
10. Faul, F., Erdfelder, E., Lang, A.-G., Buchner, A.: G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior Research Methods 39, 175–191 (2007)

Applying the *Affinto* Ontology to Develop a Text-Based Emotional Conversation System

Idoia Cearreta and Nestor Garay

Informatika Fakultatea.
University of the Basque Country/Euskal Herriko Unibertsitatea,
Manuel Lardizabal 1, 20018 Donostia, Spain
{idoia.cearreta,nestor.garay}@ehu.es

Abstract. With the recent spread of computing systems the need to enhance interactions between users and systems is evident. Conversation systems have a key role to play in achieving this. However, further efforts are needed to enhance conversation systems that use text to interact with users. This paper presents a text conversation system that includes user emotion recognition and generation, with the aim of achieving a more natural communication. The Affinto ontology is used to perform these tasks.

Keywords: Conversational System, Affective Computing, Ontology.

1 Introduction

With the spread of computing systems in recent years, there is a growing need to enhance interaction with these systems. The main problem when interacting with these systems is the need to take into account the affective values transmitted by users. In addition, a system has to make decisions based on these values and it has to transmit its own emotions. However, this is not an easy task because there are a lot of practical issues to resolve.

In this paper, a text-based emotional conversation system has been developed using the techniques employed in Affective Computing [1]. This involves the development of emotion recognition and generation processes. These processes are developed using an ontology called *Affinto* [2].

The next section reviews related work, while section 3 describes the components of the conversation system and presents an interaction example. The final section contains the conclusions that can be drawn from this work and the future work.

2 Related Work

There are a huge number of conversation systems that provide information and several services to users. In most of these systems the user has to transmit messages via speech and has to be able to receive speech signals transmitted by the systems [3]. Moreover, the majority of conversation systems found in the literature do not take into account the affective values of messages.

With respect to Affective Computing, Baldassarri et al. [4] have developed a conversation interface to enable communication between the user and a real-time

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 479–482, 2011.

virtual actor. This communication is performed using natural language and a speech synthesizer that allows the changing of parameters in order to generate affective speech. Unlike Baldassarri et al., in this paper we present a conversation system that recognizes emotions from text. The system also gives text-based affective responses. The abovementioned *Affinto* ontology is used to associate words with their corresponding affective values. Furthermore, user interactions with the system are gathered in the ontology in order to compile a correctly labelled sentence corpus for conducting further studies in the area of text-only emotion recognition.

There are several methods and techniques for recognizing emotions in texts. In our work we use the technique based on affective dictionaries [5], since the others require a large labelled corpus.

3 Prototype of an Emotional Conversation System

3.1 System Description

A global architecture has been created based on the *Affinto* ontology [2]. The main objective of this architecture is the development of multimodal affective resources. A more specific architecture has been created for the system presented in this paper. This specific architecture implements the affective interactions between the user and the system by means of text; i.e. verbal information [2]. This architecture has been useful in creating a prototype conversation system that is able to recognize a user's affective states starting from the verbal information channel.

The *Affinto* ontology represents the concepts involved in the interactions between people and systems. Moreover, it groups the instances of those concepts to infer the information required by the applications that use this ontology. *Affinto* also provides the option of representing emotions via different classification theories and the option of indicating whether a stimulus is created internally by a person or is a reaction to an external stimulus. This ontology has been created in the OWL language [6].

A method based on affective dictionaries has been used to develop the text emotion recognizer included in the conversation system. This method is based on searching the affective words of a given sentence on an existing dictionary. These dictionaries store the affective values associated with each word. The ANEW dictionary [7] has been selected for this purpose.

With regard to emotion representation, the system uses a dimensional classification incorporating the SAM (Self-Assessment-Manikin) tool [8]. SAM is a non-verbal pictorial assessment technique that consists of three scales corresponding to three dimensions: Valence, Activation and Dominance.

Depending on the emotion estimated by the recognizer, the system will create an affective response that will also be represented using the abovementioned SAM scales. However, instead of using the images of the three scales to display emotions, the system shows the values of the three dimensions integrated in a single image.

The interpreter used in the presented prototype is RebeccaAIML [9], from the ALICE Project [10]. Several changes have been made in its categories to give responses based not only on the text but also on the user emotions. These categories also incorporate affective values within the responses generated by the interpreter.

3.2 Architecture of the Text-Based Emotional Conversation System

The system able to recognize a user's affective state has several modules. Their functionality is described below:

The system starts the conversation by displaying a greeting message to the user. The user answers with a message typed on the keyboard. Then, the Affective Communication Channel module (i.e. the Verbal Channel) receives the text and sends it to the Information Extraction module. This module performs a syntactic analysis of the message to enable the Interpretation and Response Engine module to emotionally label nouns, adverbs, adjectives and verbs. This task is made possible by using the affective dictionary integrated in the *Affinto* ontology. To this end, an ontology inference engine is used to extract information about the affective values of words and those that are negation dependent are inverted. Then, the average value of the sentence is calculated in order to obtain an estimate of the user's emotion. This estimate is shown to the user on a form via a SAM image which integrates the three dimensional values. It is possible for the user to correct the values shown, if necessary. In this way, real and corrected data are stored in the ontology.

The ALICE dialogue system has been used to generate answers, changing some of the sentences to incorporate semantics about emotions. Thus, the AIML interpreter will find an appropriate affective response and will make it known via the verbal communication channel on the screen, next to the SAM image that corresponds to the system emotion.

4 Conclusions and Future Work

In this paper we have presented a prototype conversation system that uses the verbal communication channel. The method used to recognize affective values is based on the ANEW affective dictionary, by obtaining the average value of the entire sentence after emotionally labelling each word. Thus, a large corpus to classify emotions is not required. Since the message transmission is via text, this type of conversation system may be useful to people who are unable to speak or hear.

Moreover, using the *Affinto* ontology as a knowledge base that describes interactions through several modalities makes it easier to incorporate other types of affective recognition or synthesis, such as via speech. However, using this type of dictionary for text emotion recognition may have some disadvantages. For example, certain adverbs are used to emphasize or change the superlative grade of following words and thus the global affective emotional value of the sentence could change. Therefore, future work will include a more exhaustive syntactic analysis to detect these types of words.

As well as validating the *Affinto* ontology, another reason for developing this prototype was to create a small corpus based on the conversations between the users and the system. Since the users emotionally validate the sentences they transmit, it is possible to build a corpus of emotionally labelled sentences. Therefore, the corpus created could be used by a text emotion recognizer that uses other techniques.

At the present time, we are evaluating the affective text recognizer with an experiment with volunteers. In this way, we can analyze the concordances between the volunteer results and the estimates provided by the emotional recognition system.

Acknowledgments. This research work has received financial support from the Department of Education, Universities and Research of the Basque Government.

References

1. Picard, R.W.: Affective Computing. MIT Press, Cambridge (1997)
2. Cearreta, I., Garay-Vitoria, N.: Ontology for adapting interactions between humans and systems able to manage affective states. In: 18th UMAP - Workshop on User Models for Motivational Systems. Big Island of Hawaii, USA, pp. 1–12 (2010)
3. Ljunglöf, P., Larsson, S., Mühlenbock, K., Thunberg, G.: TRIK: A talking and drawing robot for children with communication disabilities. In: Proc. Nodalida 2009 - 17th Nordic Conference of Computational Linguistics. Short paper and demonstration (2009)
4. Baldassarri, S., Cerezo, E., Anaya, D.: Interacción emocional con actores virtuales a través de lenguaje natural. In: Macías, J.A., Granollers, A., Latorre, P. (eds.) Interacción'07 - VIII Congreso Internacional de Interacción Persona-Ordenador, pp. 343–352 (2007)
5. Francisco, V., Gervás, P.: Análisis de dependencias para la marcación de cuentos con emociones. Procesamiento de Lenguaje Natural 37, 1135–5948 (2006)
6. OWL (Web Ontology Language). W3C Recommendation, February 10 (2004), `http://www.w3.org/2004/OWL/`
7. Bradley, M.M., Lang, P.J.: Affective norms for English words (ANEW): Instruction manual and affective ratings. Technical Report C-1. The Center for Research in Psychophysiology, University of Florida (1999)
8. Bradley, M.M., Lang, P.J.: Measuring emotion: The self-assessment manikin and the semantic differential. Journal of Behavior Therapy and Experimental Psychiatry 25, 49–59 (1994)
9. RebeccaAIML. Recuperado el 5 de Abril de (2011), `http://rebecca-aiml.sourceforge.net/`
10. ALICE AI Foundation. Recuperado el 8 de Septiembre de (2010), `http://www.alicebot.org/about.html`

Augmented Mirror: Interactive Augmented Reality System Based on Kinect

Lucía Vera, Jesús Gimeno, Inmaculada Coma, and Marcos Fernández

Instituto de Robótica, Universitat de València
{Lucia.Vera,Jesus.Gimeno,Inmaculada.Coma,Marcos.Fernandez}@uv.es

Abstract. In this paper we present a virtual character controlled by an actor in real time, who talks with an audience through an augmented mirror. The application, which integrates video images, the avatar and other virtual objects within an Augmented Reality system, has been implemented using a mixture of technologies: two kinect systems for motion capture, depth map and real images, a gyroscope to detect head movements, and control algorithms to manage avatar emotions.

Keywords: Augmented Reality, Motion Capture, Virtual Characters.

1 Introduction

In the field of HCI, combining the real world with virtual information and introducing virtual characters interacting with a real audience are elements that add richness and complexity to presentations and public events.

In this context we have developed an Augmented Reality (AR) application that allows the audience to interact and talk with a virtual character through a large screen called augmented mirror. The avatar movements are controlled by an actor in real-time using a Motion Capture (MoCap) system. That avatar, together with video images from the real world and other virtual objects, are integrated in a system that also incorporates the voice of the character and the manipulation of virtual objects.

In a review of the literature we find different MoCap systems used for the animation of virtual characters. The MoCap devices use different technologies to acquire positions. Thus, in Waldo C., a virtual puppet, a *mechanical system* was used to track body joint angles using an exo-skeleton structure [1]. Other systems use *electromagnetic sensors* [2], or *optical systems* with markers and cameras. Finally, some use optical systems without markers using an image recognition algorithm to detect positions and movements [3].

Within AR applications, MoCap devices are also used to control the interaction between the user and the virtual objects. Sometimes optical systems with markers [4] are used to detect positions but, in other cases, more complex systems are utilized that are based on ultrasounds or inertial sensors that detects movements in a wide range of use [5, 6].

Aiming to achieve a system that captures the actor's movement in terms of limited space and lighting conditions, including facial expressions and lip movements, we used a combination of technologies: a Kinect device for MoCap, a gyroscope for head

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 483–486, 2011.

movements, and a WiiMote with control algorithms for facial expressions. Moreover, it has been necessary to create an augmented video image blending the avatar and the real environment, using another Kinect depth map for that purpose. However, Kinect presents some restrictions that we have solved with different solutions described in the next section.

2 System Description

The proposed system is composed of two main parts, called scenarios (see Figure 1). In the first one, the control scenario, the actor tracking is performed using a MoCap system. In the second one, the augmented scenario, the audience can interact with the avatar and other virtual objects. These two scenarios are connected using a Client-Server model, via an Ethernet connection, so that the avatar can be remotely controlled.

With regards to the control scenario, a number of input devices are combined into an enhanced MoCap system (see Figure 2) which serves the actor to control the avatar movements, facial expressions, lips movement and also some predefined virtual objects. Moreover, input data from these devices is received and processed by a server application to control the augmented scenario (the avatar and the virtual objects).

The main MoCap device is a Kinect camera that controls the avatar movement. By using OpenNI (Open Natural Interface) SDK, Kinect performs actor tracking in real time, with an 11 bones avatar structure (see Figure 1). In our system, this control was not enough to obtain a realistic avatar and, therefore the tracking was enhanced in 4 different ways: head orientation, lips movement while talking, facial expressions and automatic gestures (blink, hands, feet, etc.).

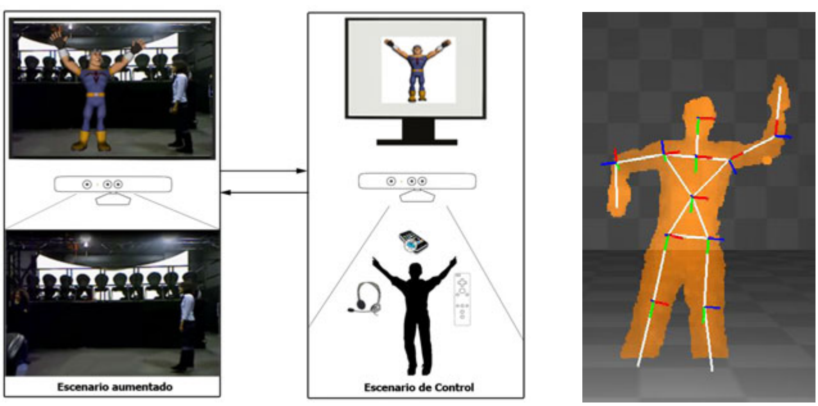


Fig. 1. General system structure (left); Kinect avatar skeleton (right)

Head orientation is estimated using a mobile phone attached to a cap. This can be achieved with an android application that computes heading, pitch and roll from the phone sensors data (accelerometer and compass) and send it to the server application through a Wi-Fi connection. This global information, referenced to the earth gravity and North Pole, is computed to calculate head orientation according to the actor orientation. For lips movement while talking with the audience, the actor is provided with a wireless microphone. An amplitude based algorithm has been developed to calculate, lips aperture values (horizontal and vertical), which simulates the real mouth movement while talking. A WiiMote is used to control avatar facial expressions and 3D virtual objects. The actor can easily choose the avatar facial expression, using the WiiMote buttons, from five different predefined emotions. In the same way, 5 virtual animated objects can be shown or hidden. Finally, other avatar movements are computed automatically. Blinking velocity is controlled depending on the selected emotion; hand and finger movements are calculated from the arm movement. All this information is sent to the augmented scenario in real time, making possible a live interaction between the real actor and an audience.

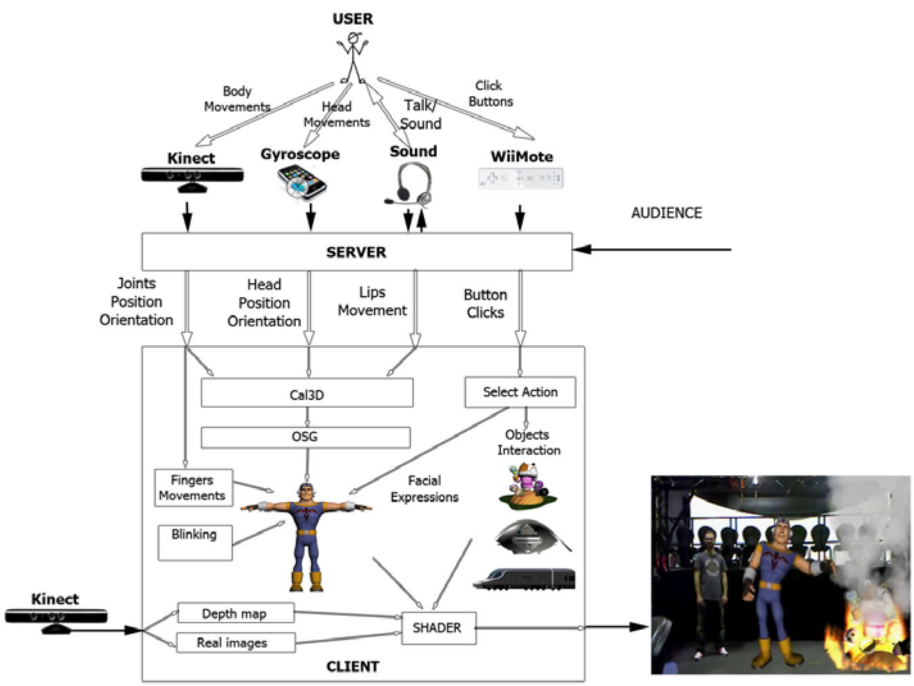


Fig. 2. Scheme of the overall application

The augmented scenario is where the audience interacts with the avatar and the virtual objects via the augmented mirror. The audience can talk with the avatar or walk around him and the virtual objects while they can see themselves in a large LED screen (Figure 4). An important element of our augmented visualization is the correct occlusion between real and virtual information. This augmented images displayed on a 4x3 meter LED screen provide a more immersive multiuser augmented reality experience. A Kinect camera is used to capture the real image of the audience which is used in both scenarios. In the control scenario the real actor can see the audience to interact with them, while this image is merged with the avatar and the virtual elements to generate the augmented mirror image. The occlusion handling is implemented on a

GPU shader, which takes the real image and the depth information captured by Kinect, mixing it with the virtual scene. The implementation of the augmented visualization is based on OpenSceneGraph, Cal3D and our own developed libraries. The avatar skeleton (see Figure 3), has been designed to perform all the gestures and facial expressions desired.

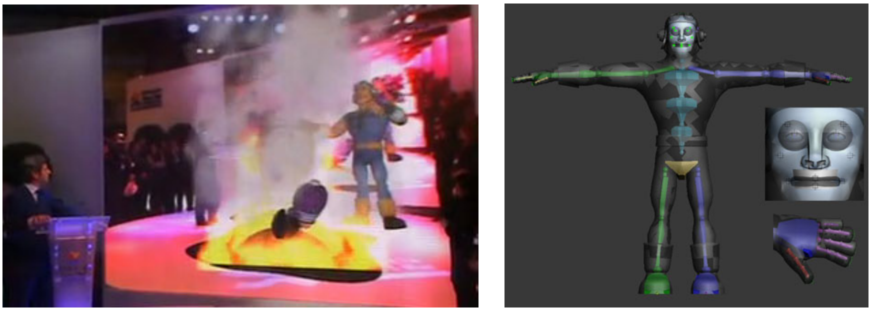


Fig. 3. Augmented mirror (left); avatar skeleton (right)

3 Conclusions

This paper has presented an augmented mirror where a virtual actor interacts with an audience. Kinect motion capture capabilities have been used and enhanced with other 3 input devices to perform a MoCap system which allows the actor to control the avatar movements, facial expressions, lips movement and also some predefined virtual objects. The system has been successfully used in a real marketing presentation involving touristic contents (FITUR'11), where nearly 100 person enjoyed the augmented mirror experience.

References

1. Walters, G.: The story of Waldo C. Graphic. Course Notes: 3D Character Animation by Computer. In: ACM SIGGRAPH 1989, Boston (1989)
2. Huang, Y., Kallman, M.: Interactive Demonstration of Pointing Gestures for Virtual Trainers. In: Jacko, J.A. (ed.) HCI International 2009. LNCS, vol. 5611, pp. 178–187. Springer, Heidelberg (2009)
3. Wren, R., Azarbayejani, A., Darrell, T., Pentland, A.P.: Pfinder: real-time tracking of the human body. Trans. Pattern Anal. Mach. Intelligence 19(7), 780–785 (1997)
4. Dorfmüller, K.: Robust tracking for augmented reality using retroreflective markers. Computer & Graphics 23(6), 795–800 (1999)
5. Foxlin, E., Harrington, M., Pfeifer, G.,, C.: A Wide-Range Wireless Tracking System for Augmented Reality and virtual set applications. In: Proc. of SIGGRAPH 1998 (1998)
6. Vlasic, D., et al.: Practical motion capture in everyday surroundings. ACM Transactions on Graphics 26(3) (2007)

Calls for Interaction: The More the Better? User Experience of 3D Carousel and Additional Interaction Techniques

S. Shyam Sundar[1,2], Saraswathi Bellur[1], Jeeyun Oh[1], and Haiyan Jia[1]

[1] Media Effects Research Laboratory,
The Pennsylvania State University, USA
[2] Department of Interaction Science,
Sungkyunkwan University, Seoul, South Korea
{sss12,saras,jzo120,hjia}@psu.edu

Abstract. We perform a user study to investigate the psychological consequences of adding interaction techniques to the interface. In a between-subjects experiment (N = 143), we explore how (i) variations in sheer number of interaction techniques and (ii) addition of a novel technique, i.e., 3D carousel, influence the volume of users' actions, their memory, perceptions of interactivity, as well as their attitudes and behaviors toward a website. Power usage is examined as a potential moderator. First-cut findings from self-reports and log data indicate that the 3D carousel feature has a strong impact on user experience, both positive and negative. It also moderates the curvilinear effect of adding traditional interaction techniques to the interface.

Keywords: Interaction Techniques, 3D carousel, user experience, user engagement.

1 Introduction

The range and richness of interaction techniques on websites offer users unprecedented variety and choice in interacting with an interface. For instance, the 3D carousel is fast becoming a common feature on websites, joining a growing cadre of interaction tools. However, little is known about the user experience created by these techniques. Does the increased number of calls for interaction necessarily enhance user experience? How does this feature combine with other techniques on the interface to influence higher-order outcomes? With this in mind, the purpose of our study is twofold: (1) to examine the effect of the addition of interaction techniques to user experience, their memory, attitudes and behaviors, and (2) to specifically explore the effect of the *3D carousel* technique, and how it affects user experience?

2 Effects of Interaction Techniques

We define interaction techniques as any feature embedded in the website that offers various modes of interactions to users, based on the tripartite model of interactivity effects proposed by Sundar [1], which distinguishes between *modality interactivity,*

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 487–490, 2011.

source interactivity (customization) and *message* interactivity (contingent exchanges of messages). Interaction techniques come under the first of the three kinds of interactivity, in that they afford various modes of accessing information on an interface. Different types of interaction techniques add to the interactivity potential of the interface. However, literature suggests that the effects of interactivity are not always linear or positive. Empirical observations have revealed a so-called "interactivity paradox" [2], with studies pointing to an optimal threshold beyond which the effects of interactivity diminish [3]. Just what, how or when exactly is such a threshold reached remains an open question that is worthy of investigation. Aside from increases in the availability of interaction techniques, the presence of novel techniques such as the 3D carousel, with its characteristically revolving motion, invite active user action (e.g., flipping or swiping) that add to its interactivity potential. Its constantly rotating feature makes the 3D carousel playful, engaging and capable of providing rich forms of interaction, as claimed by Björkskog et al [4]. However, such enthusiasm for it is yet to be validated empirically with users, especially when it co-exists with other interaction techniques on the interface. Therefore, we experimentally examine how the addition of 3D carousel, combined with other interaction techniques, affects user experience and the degree of user engagement with a Web interface.

3 Method

We conducted a 4 (Click only, Click + Slider, Click + Slider + Drag, Click + Slider + Drag + Mouseover) × 2 (presence or absence of 3D carousel) between-subjects fully-crossed factorial experiment. Participants recruited from undergraduate classes (N = 143; 93 females) were randomly assigned to one of the eight conditions. Eight prototype websites were especially constructed for this experiment based on an online exhibition entitled "Guitar/Bass Timeline" [5]. Except for the number of interaction techniques employed, all eight versions of the prototype website shared the same content and page layout. We embedded the manipulation of 3D carousel in the homepage (Figure 1). Whenever participants clicked on a guitar in the homepage, the site played a brief audio-clip of a classic riff produced by that particular guitar/bass.

Fig. 1. Homepage with 3D carousel vs. without 3D carousel

In both conditions, clicking on one of the guitars took participants to the next layer of site content, which featured a slider (for going across a timeline), a drag feature (allowing users to drag a guitar-pick on the instrument for more information), and/or a mouseover function (for accessing embedded textual information in so-called "hotspots"). Examples of stimuli are accessible at <http://tinyurl.com/3f8ypmk>. The websites for all eight conditions were elaborately pretested for usability. Based on literature which identifies browsing as the quintessential interaction task [6], we instructed our study participants to browse the stimulus website fully and completely. To make the task as realistic as possible, no time constraints were imposed, nor were the participants told that their memory for site's content would be tested later in the study. All questions were measured on a 7-point Likert scale, except for recognition and recall memory measured through multiple-choice and open-ended questions. *Power usage* was measured via 12 items about the degree of participants' liking of technology and their extent of use [7] (α =.83). *Perceived interactivity* was measured by 3 items (for e.g., the degree to which it allows them to perform lots of actions) (α = .76) based on [8]. *Perceived ease of use* was measured by 3 items (α = .74) adapted from [9]. Behavioral intention toward website was measured by 6 questions (α = .93) [10]. We collected log data using jQuery to measure user actions (average frequency of page visits and average number of user interactions per hotspot).

4 Results

An interaction effect showed that the addition of interaction techniques to the interface served to enhance user **perceptions of interactivity** only in the absence of 3D carousel, F (3, 134) = 4.11, $p < .01$. The relationship between the number of interaction techniques and **perceived ease of use** assumed an inverted-U shape when the 3D carousel feature was present on the interface, F (3, 134) = 4.39, $p < .01$ (see Figure 2). The analysis of **behavioral intentions** (e.g., forwarding and recommending the site to others) yielded a similar inverted-U interaction, F (3,134) = 3.01, $p < .05$ (Figure 3).

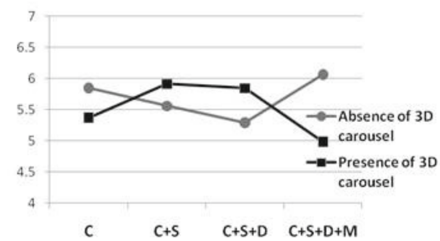


Fig. 2. Perceived ease of use
(C: Click, S: Slide, D: Drag, M: Mouseover)

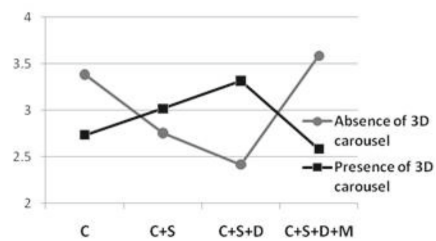


Fig. 3. Behavioral intentions
(C: Click, S: Slide, D: Drag, M: Mouseover)

A significant main effect revealed that 3D carousel served to degrade **aural recognition memory** of the guitar riffs, F (1, 134) = 6.17, $p < .05$. In terms of **user actions,** users were most likely to interact with information hotspots when 3D carousel was present, F (1, 131) = 11.62, $p < .001$. However, users tended to visit the homepage (i.e., the main page with the guitars) less frequently when 3D carousel was present, F (1, 132) = 3.98, $p < .05$. **Power usage** positively predicted perceptions of interactivity, F (1, 134) = 8.14, $p < .01$, and behavioral intentions toward the site, F (1, 134) = 3.95, $p < .05$. As a moderator, power usage was positively associated with recall memory for content in the 3rd and 4th conditions, but inversely related to recall when only two basic interaction techniques were present, F (3, 135) = 3.68, $p <. 05$.

5 Discussion

Taken together, the 3D carousel boosted perceptions of interactivity, so much so that it overrode common interaction techniques such as click, slider, and drag. It also stimulated user interaction. However, this appeal of 3D carousel came at a cognitive cost, degrading users' memory for audio that accompanied the guitars and inhibiting their tendency to visit the homepage. It negatively affected perceived ease of use and behavioral intentions when the study website was saturated with four interaction techniques, even when users reported that a site with 3D carousel was easier to use and were likely to recommend it. This parallels previous findings [11] that the 3D carousel is more suitable for aesthetic appeal than task efficiency. In addition, more interaction techniques aided the information uptake for power users, encouraging us to consider individual differences relating to technology competency when predicting user experience with new as well as multiple combinations of interaction techniques.

Acknowledgments. This research is supported by the U.S. National Science Foundation under Grant No. IIS-0916944 and the Korea Science and Engineering Foundation under World Class University (WCU) Grant No. R31-2008-000-10062-0.

References

1. Sundar, S.S.: Social psychology of interactivity in human-Website interaction. In: Joinson, A.N., McKenna, K.Y.A., Postmes, T., Reips, U.-D. (eds.) The Oxford Handbook of Internet Psychology, pp. 89–104. Oxford University Press, Oxford (2007)
2. Bucy, E.P.: The debate. The Information Society 20(5), 371 (2004)
3. Sundar, S.S., Kalyanaraman, S., Brown, J.: Explicating web site interactivity: Impression formation effects in political campaign sites. Communication Research 30, 30–59 (2003)
4. Björkskog, C., Jacucci, G., Lorentin, B., Gamberini, L.: Mobile implementation of a web 3D carousel with touch input. In: Proc. of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services, pp. 1–4. ACM, New York (2009)
5. http://www.empsfm.org/flash/guitbass/index.html
6. Zhai, S., Smith, B.A., Selker, T.: Improving Browsing Performance: A study of four input devices for scrolling and pointing tasks. In: Proc. of the 6th IFIP Conference on Human-Computer Interaction, pp. 286–292 (1997)
7. Marathe, S.S., Sundar, S.S., Bijvank, M.N., van Vugt, H., Veldhuis, J.: Who are these power users anyway? Building a psychological profile. Paper presented at the 57th Annual Conference of the International Communication Association (2007)
8. Kalyanaraman, S., Sundar, S.S.: The psychological appeal of personalized online content in Web portals: Does customization affect attitudes and behavior? Journal of Communication 56, 110–132 (2006)
9. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Quarterly 13(3), 319–339 (1989)
10. Hu, Y., Sundar, S.S.: Effects of online health sources on credibility and behavioral intentions. Communication Research 37, 105–132 (2010)
11. Wang, S., Poturalski, M., Vronay, D.: Designing a generalized 3D carousel view. In: CHI 2005 Extended Abstracts on Human Factors in Computing Systems, pp. 2017–2020. ACM Press, New York (2005)

Can Persona Facilitate Ideation? A Comparative Study on Effects of Personas in Brainstorming

Xiantao Chen[1], Ying Liu[2], Ning Liu[3], and Xiaojie Wang[4]

[1] Beijing University of Posts and Telecommunications, Nokia Research Center Beijing, 100176 Beijing, China
ext-xiantao.chen@nokia.com
[2] Nokia Research Center Beijing, 100176 Beijing, China
ying.y.liu@nokia.com
[3] School of Psychology, Shandong Normal University, 250014 Jinan, China
sdliuning@gmail.com
[4] Beijing University of Posts and Telecommunications, 100876 Beijing, China
xjwang@bupt.edu.cn

Abstract. Personas are results from user studies and viewed as a design and a communication tool in user-centered design processes. There were many studies addressing how to create good personas but what types of personas and how personas could help in ideation processes were less discussed in past works. In this paper, we conducted a comparative study to explore effects of personas on the ideation process and idea qualities in a brainstorming setting. The results indicated that personas could enhance the ideation process and design deliverables on two aspects: personas could help both individual designer and a group of designers focus on the target user group during the ideation process; and the delivered ideas or concepts were viewed more relevant to the user groups and were more comprehensive.

Keywords: Comparative study, Persona, Ideation, Brainstorming.

1 Introduction

Persona is a method introduced by Cooper to describe user profiles [1]. A well-cited definition of it is that personas are fictional archetypical characters that represent distinct groupings of behavior, goals and motivations identified in a study [2]. There were two different applications of personas in existing design practices: as a communication tool or a design tool [3]. However, little research has been done to prove the benefits of persona as a design tool.

In this research we aimed to investigate whether persona can facilitate ideation as a design tool in the early stage of a user-centered design (UCD) process. We first built three personas for middle school students in top cities of China by following Nielsen's ten steps to personas [4]. The user data that our personas were built on were from the Youth Study project. The purpose of the project was to understand their lifestyle and usage of information and communication technologies (ICTs) of urban youth in China. Based on the three personas, we conducted an exploration on effects of persona in the early concept design phases in the UCD process. Design students

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 491–494, 2011.

were invited to a brainstorming session and they were either presented with personas or not. We observed and coded the design process and idea qualities from the brainstorming sessions. Our results found that proper personas can help designers to deliver ideas that are more appropriate for the target user group. The study and results are presented in the rest of the paper.

2 A Comparative Experimental Study

Fourteen participants took part in the brainstorming session and they were divided to two groups with respectively seven people. All of them were undergraduates in industrial design and information design in a top university. For Group A, we present both three personas and raw materials including user study transcriptions and dialogues during their brainstorming session and for Group B, only raw materials were presented. The assigned topic was "Mobile solutions for senior middle school students" for both groups. And two relevant tasks were assigned at the same time. In the first task, participants were given an hour for reviewing the user study materials and generate ideas individually. In the second task, participants were asked to join a cooperative ideation session for one hour. Two final concepts needed to be sketched and delivered by each group.

Shah [5] proposed two approaches for evaluating the effectiveness of ideation: process-based measures that measure the process of ideation, and outcome-based measures relating to the results. During the experiment, all ideas created in both design tasks were recorded and analyzed. Based on the idea evaluation framework introduced by Barki [6], ten design experts were invited to evaluate ideas based on an 11-point Likert scale on a few aspects: novelty, uniqueness, perfectness, value, relevance with user needs, suitability for the users and an overall score. Since the ideas produced in the first task were quick ones with little elaborated descriptions, only novelty, uniqueness, relevance with the user group and suitability for the users of them were evaluated. About the process of brainstorming sessions, observations was conducted to track when the first idea was generated and the usage of materials.

3 Results

3.1 Ideation Process

We analyzed our observation notes in both groups. From the results, we found that the group A with personas spent less time in producing the first idea than the group B without personas. All participants in Group A generated their first ideas in the first half hour, however only two participants in Group B generated their first ideas. About the usage of materials, all participants in Group A shifted the usage of personas and raw materials to generate ideas. They thought that both of them were important to inspire novel ideas. Typical usages of personas included that they were used to get overview to target users, inspire novel ideas from the features and refine final concepts. Participants in Group A reported that details and distinct features in personas were important to inspire ideas. In Group B, several participants complained that twelve pages' raw materials were too long to read. Some of them gave up reviewing raw materials and generated ideas form their own experiences.

Table 1. The t-test of idea quality in first task between Group A and Group B

Category		Group A	Group B	*t*	*p*
Is it novel	*M*	6.09	5.57	1.50	0.159
	SD	0.79	0.46		
Is it unique	*M*	5.97	5.47	1.49	0.162
	SD	0.74	0.48		
Is it relevant with user need	*M*	7.13	6.10	3.49	0.004*
	SD	0.54	0.57		
Is it suitable for user	*M*	6.74	5.81	3.89	0.002*
	SD	0.37	0.51		
Overall score	*M*	6.41	5.75	2.24	0.045*
	SD	0.64	0.42		

3.2 Idea Quality

In the first task, participants in Group A generated an average of 6.29 new ideas per person and an average of 8.00 new ideas in Group B per person. Although Group A produced less ideas than Group B, there were significant difference on quality of ideas between the two groups according to the expert evaluation results. The t-test results indicated that idea quality of Group A is better than that of Group B on a few characteristics: relevance with user (t (12) =3.49, p<0.005), suitability for user (t (12) = 3.89, p<0.005) and overall scores (t (12) = 2.24, p<0.05) (see Table 1). The significant differences revealed that participants in Group B inclined to produce ideas that were not suitable for the design topic. No significant difference was found on the qualities of "novelty" (t (12) = 1.50, ns) and "uniqueness" for ideas from both groups (t (12) = 1.49, ns). The results from the first design task showed that engaging personas in individual ideation could help participants focus on the assigned topic.

Table 2. The t-test of idea quality in second task between Group A and Group B

Category		Group A	Group B	*t*	*p*
Is it novel	*M*	6.67	5.94	1.92	0.073
	SD	0.71	0.88		
Is it unique	*M*	6.39	5.67	1.62	0.126
	SD	0.96	0.94		
Is it perfect	*M*	7.17	5.33	3.67	0.002*
	SD	0.71	1.32		
Is it valuable	*M*	6.67	5.94	2.14	0.048*
	SD	0.66	0.77		
Is it relevant with user need	*M*	7.00	6.06	2.20	0.042*
	SD	0.79	1.01		
Is it suitable for user	*M*	6.67	5.83	2.24	0.040*
	SD	0.71	0.87		
Overall score	*M*	6.94	5.53	3.84	0.001*
	SD	0.68	0.87		

In the second design task, the two final concepts contributed in each group were evaluated. The quality of the concepts from Group A was ranked higher than that from Group B: significant difference were found on perfectiveness ($t(12) = 3.67$, $p < 0.005$), value ($t(12) = 2.14$, $p < 0.05$), relevance with users ($t(12) = 2.20$, $p < 0.05$), suitability for users ($t(12) = 2.24$, $p < 0.05$) and the overall scores ($t(12) = 3.84$, $p < 0.005$) (see Table 2). The results showed that personas could facilitate the group ideation session and enable designers create complete concepts.

4 Conclusion and Further Work

In this research, we explored the usage of personas in early ideation processes. A comparative study proved that personas enhanced the effectiveness of brainstorming on two aspects: it helped designers to stay on the assigned design topic during the individual brainstorming session and it also helped a group of designers to agree on the topic for the group concept and delivered comparatively complete concept proposals. Moreover, personas could also potentially improve idea qualities especially on measures to indicate relevance to target users.

Although benefits of personas in brainstorming sessions were discovered, limitations of our personas have been identified too. Some designers complained our personas were lack of detailed information and they had to switch between personas and with the raw materials. How to create comprehensive personas with detailed information of users for ideation processes need to be considered in our further work.

References

1. Cooper, A.: The Inmates Are Running the Asylum, pp. 123–124. SAMS/Macmillan, Indianapolis, IA (1999)
2. Calde, S., Goodwin, K., Reimann, R.S.: Orcas: The First Integrated Information System for Long-term Healthcare Facility Management. In: Dykstra-Erickson, E. (ed.) Proceedings of Conference on Human Factors in Computing Systems, Minneapolis, Minnesota, USA, pp. 2–16. ACM Press, New York (2002)
3. Guojonsdottir, R., Lindquist, S.: Personas and Scenarios: Design Tool or A Communication Device. In: Proceedings of the 8th International Conference on the Design of Cooperative Systems, pp. 165–176 (2008)
4. Nielsen, L.: Engaging Personas and Narrative Scenarios. Vol. 17, PhD Series. Samfundslitteratur, Copenhagen, pp. 1–331 (2004)
5. Shah, J., Vargas-Hernandez, N.: Metrics for Measuring Ideation Effectiveness. Design Studies 24(2), 111–134 (2003)
6. Barki, H., Pinsonneault, A.: Small Group Brainstorming and Idea Quality: Is Electronic Brainstorming the Most Effective Approach? Small Group Research 32(2), 158–205 (2001)

Children with Special Needs: Comparing Tactile and Tangible Interaction

César Ortea Suárez, Javier Marco, Sandra Baldassarri, and Eva Cerezo

Computer Science Department,
Instituto de Investigación en Ingeniería de Aragón,
Universidad de Zaragoza
{622384,javi.marco,sandra,ecerezo}@unizar.es

Abstract. In this paper a comparison of the same computer game with two interaction styles is achieved: through tactile interaction in a digital board or using tangible interaction on a tabletop. Tests were carried out with children with special needs, who have different degrees of disability. The aim of the paper is to compare usability and accessibility of each interface, as well as the experiences of children playing with them. Preliminary results indicate the necessity to provide feedback continuously, both hearing and visual, in order to facilitate the understanding of each task and its progression. In addition, the outcome obtained from a questionnaire show a significant preference for the tangible version of the game.

Keywords: Tactile, tangible, digital board, tabletop, special education, game.

1 Introduction

Computer games are very valuable education-learning tools for children, because they allow users to learn through active participation, to promote problem solving and to emphasize exploration and self-discovery. Also, games are an entertainment for children without perceiving that, at the same time, they are learning.

In recent years, children have gained role as users of interactive technologies. Consequently, there has been a growing interest in adapting and validating different evaluation methods to the cognitive and social development of children [1]. Also, depending on the children's disability, it is necessary to adapt the system's interaction style to their capabilities. Graphical User interfaces (GUIs) based on mouse and keyboard interaction are the most used today, although more natural and physical interaction models are becoming increasingly popular [2]. Natural interaction methods, including tactile interaction and tangible user interfaces (TUIs), have been explored as a way of bringing children to the digital contents. In particular, TUIs combine physical and virtual objects in interactive and physical environments [3] and, therefore, they are more exploratory, collaborative and expressive [4].

In this paper, we carry out a comparison of the same game with two kinds of natural interaction: tactile interaction in digital board and tangible interaction in horizontal tabletop. Children who participate in the evaluation tests belong to special education school in order to learn how to fit these two interaction styles to the needs

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 495–498, 2011.

of children with cognitive problems [5] [6] [7]. Consequently, the evaluation methods have to be chosen and adapted according to their capacities, with the purpose of obtaining information about usability, accessibility and user experience (UX) of both interfaces.

2 The Farm Game with Different Interaction Styles

We compare the same video game for children in two different devices: horizontal tabletop based on tangible interaction (Fig. 1 left) and a digital board based on tactile interaction (Fig. 1 right), as both enable small groups of children playing together.

Fig. 1. Tangible (left) and tactile (right) interfaces

The **tangible interface** is composed of a horizontal tabletop (NIKVision) specially designed for children [8]. The interaction with the games is through physical manipulation of toys on the table. The device has an active output image on the tabletop and on a monitor next to it.

On the other hand, the **tactile interaction** on the digital board is achieved by pressing on an image which is projected onto a vertical screen (TouchIT model). Also, the board is situated near to the ground in order to children can touch easily everywhere.

For both interactive devices, it has been designed the same game: a Farm game, which consists of a final goal, several sub-goals, characters controlled by the player, an autonomous characters and interactive objects. The ultimate goal of the game is to help the farmer to make a cake for his son's birthday. To achieve this goal, children must complete three sub-goals: get three strawberries, four eggs and a bucket of milk.

The players can control several characters: a hen, a cow, a sheep and a pig. Strawberries can be picked up with any animal, while eggs and milk can only be obtained with the hen and cow, respectively. The farmer is an autonomous character, which gives hearing feedback (instructions about who, where and how pick up the ingredients). In addition, there is different visual feedback in both versions:

- Tabletop: The projection on the table indicates where to place the animals, where strawberries are and how many eggs or milk have been achieved.
- Digital board: At the top of the screen, a box indicates the object where the animal must be placed. At the bottom, several boxes indicate the amount of strawberries, eggs or milk already picked up and the amount that is still needed.

Finally, the interactive objects are: plants (with or without strawberries), the nest and the bucket. Every animal can be moved or placed on these virtual objects and it can perform an action on them, giving back feedback about task's progress (farmer's instructions).

The sequence of actions to perform in order to complete a task varies in each interaction style:

- Tangible: first, a toy animal is placed on a virtual object on the table surface and then, the toy can be shaken in the plants to get a strawberry, or jumps can be done with the hen and cow toys to lay an egg in the nest or to give milk in the bucket respectively.
- Tactile: it requires a sequence of touches on the board to complete an action: first, an animal is touched to activate it, then an object is touched and the animal moves to it and, finally, the object is touched again to trigger an action.

3 Comparative Test

An evaluation session has been carried out involving special education children who have different degrees of cognitive and/or physical disabilities. In particular, eight children with ages between 6 and 11 years, participate in couples.

As has been said before, the evaluation methods used are intended to get usability, accessibility and UX problems in both interaction styles. Specifically, the evaluation methods selected for these test sessions, are the following: usability test (evaluators write observations relating to children's interaction with both interfaces), video analysis, LOG (automatic record of events and actions that happen in the game), likeability questionnaire (to obtain the preference of children) and SEEM questionnaire (Structured Expert Evaluation Method) in order to capture experts' view of the interaction problems of the children.

We present the preliminary outcomes and results from our observations and questionnaires, but a more thorough analysis from the videos and LOGs must be still carried out.

The observation reveals that children interact following the instructions in the same order they receive them (data obtained from LOGs).

The results of the questionnaires mainly address to the need to provide more feedback, both visual and hearing. In particular, in the tangible version of the game, it is necessary to provide back visual or hearing feedback to show children how many objects are still needed. Also the major goal of the game (to make a cake) must be clearly represented so that children can understand the aim. The tactile version of the game, on the other hand, has to provide more visual feedback, especially on those positions where children have to trigger an action. In addition, farmer's instructions must be given continuously and must be as simple as possible.

A widespread problem is the difficulty for children to complete the task of picking up strawberries, because they don't perceive clearly where the strawberries are placed. Therefore, it is necessary to give back visual feedback in these plants.

Finally, regarding to the answers given in the likeability questionnaire, 5 out of 8 children prefer tangible version and only one person prefer the tactile version (2 children didn't say anything).

In summary, this study shows us the differences between both versions and the importance of giving back visual and hearing feedback so that children can understand the game's progress. In tactile version, hearing feedback must be simplified and visual feedback must be displayed in the interactive object (not in other place), while in tangible version feedback should be given to show how many actions are in the task. Regarding to the answers of the likeability questionnaire, 5 out of 8 children prefer tangible version and only one person prefer tactile version (2 children didn't say anything).

Acknowledgements. This work has been partially financed by the University of Zaragoza, through the Project called "AVIM-Agentes Virtuales Inteligentes y Multimodales" and the Government of Aragon through Cooperation projects between University Departments and Secondary Schools.The authors thank the collaboration of CPEE Alborada School.

References

1. Baauw, E., Markopoulos, P.: A comparison of think-aloud and post-task interview for usability testing with children. In: Proceedings of the 2004 Conference on Interaction Design and Children: Building a Community, IDC 2004, pp. 115–116 (2004)
2. Xu, D.: Design and Evaluation of Tangible Interfaces for Primary School Children. In: Proceedings of the 6th International Conference on Interaction Design and Children, IDC 2007, pp. 209–212 (2007)
3. Fails, J.A., Druin, A., Guha, M.L., Chipman, G., Simms, S., Churaman, W.: Child's Play: comparison of Desktop and Physical Interactive Environments. In: Proceedings of Interaction Design and Children (IDC 2005), pp. 48–55 (2005)
4. O'Malley, C., Fraser, D.S.: Literature Review in Learning with Tangible Technologies. NESTA Futurelab, 12 (2004)
5. Li, Y., Fontijn, W., Markopoulos, P.: Tangible tabletop game supporting therapy of children with Cerebral Palsy, pp. 182–193. Springer, Heidelberg (2008)
6. Piper, A.M., O'Brien, E., Morris, M.R., Winograd, T.: SIDES: a cooperative tabletop computer game for social skills development. In: Proceedings of the 2006 20th Anniversary Conference on Computer Supported Cooperative Work (CSCW 2006), pp. 1–10. ACM, New York (2006)
7. Baldassarri, S., Cerezo, E., Blasco, G.: Juegos educativos configurables para Educación Especial. In: En: XI Congreso Internacional de Interacción Persona Ordenador, pp. 99–108 (2010) (in Spanish)
8. Marco, J., Baldassarri, S., Cerezo, E.: Bridging the Gap between Children and Tabletop Designers. In: Proceedings of the 9th International Conference on Interaction Design and Children, IDC 2010, pp. 98–107 (2010)

Coupling Interaction and Physiological Metrics for Interaction Adaptation

Luís Duarte and Luís Carriço

LaSIGE & Department of Informatics, University of Lisbon
Edifício C6, Campo-Grande, 1749-016 Lisboa
lduarte@lasige.di.fc.ul.pt, lmc@di.fc.ul.pt

Abstract. We present an adaptation system whose goal is to provide users with interaction experiences tailored to their current physiological status and performance. The system captures emotion, motion and application related metrics to proactively adjust the available interaction patterns. Interacting in different environments – stationary/mobile – or under different emotional status – relaxed/stressed– can affect performance, engagement and enjoyment. This contribution describes the initial design steps in the creation of an interaction adaptation engine.

Keywords: Physiological Signals, Adaptation, User Performance.

1 Motivation

Physiological user interfaces are typically employed to perceive real body signals from users, while utilizing that information during the interaction period [7]. These interfaces are mainly used to determine motion [2] (through muscle tension detection, video capture or accelerometers) or emotion [6] (through skin conductivity or heart rate variance). Human beings are capable of expressing both observable (e.g. blushing, frowning, etc.) and concealed (e.g. increasing heart beat rate, altering body temperature, etc.) reactions to certain events. While analysts and the most recurrently used mechanisms are able to retrieve a significant amount of data with traditional assessment techniques (e.g. camera recording, ethnography, questionnaires, etc.), these are considered highly subjective and possess an interesting negative effect for the former [3]: they typically fail to address complex interaction patterns. On the other hand, Rowe [5] states that some of the advantages of these interfaces are linked with being multi-dimensional, proving to be capable of providing alternative views on different issues. Lastly, physiological signals are typically continuously gathered, enabling a faster and more accurate detection of emotional or workload shifts.

However, one domain in which physiological interaction mechanisms have failed to prosper in is mobile environments. There are multiple examples of studies on the use of such interaction mechanisms in desktop or stationary settings in a diversity of areas, ranging from entertainment to healthcare. The most recent examples of the use of physiological mechanisms in mobile settings stem from the NikeRunning [4] program and from a system which collects physiological data from users with the aim of improving positive emotions in their daily lives [8]. Nevertheless, neither of these

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 499–502, 2011.

approaches addresses some of the most important challenges of mobile environments: constant transitions to/from different settings, interaction with peers, the device's characteristics, to cite a few.

Taking these factors into account and based on a set of experiments we have been performing to assess user behavior and interaction patterns using physiological indicators [1], we envisioned an adaptation engine for both stationary and mobile settings, which combines application log data with physiological measures to better tailor interaction styles and user interfaces to the operators' current status.

2 Adaptation Engine

The adaptation engine encompasses the use of physiological assessment mechanisms to evaluate user performance. Our approach comprises the following components:

- **Emotional Activity Module:** is related with shifts in determined biological signals, such as heart beat rate or skin conductance. Since these are two of the most researched metrics, the purpose is to identify periods of increased stress or anxiety. With this module we intend on providing the necessary mechanisms to pro-actively improve user experience and provide users with a balanced interaction.

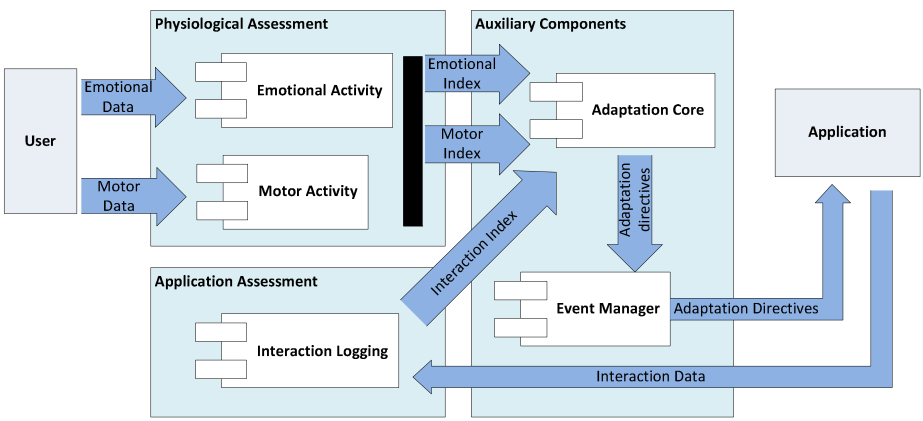


Fig. 1. Adaptation Engine's architecture

- **Motor Activity Module:** is concerned with the identification of user movement. In particular, it identifies whether a user is walking, if he/she waved his/her arms, etc. The coupling of such module with the emotional one allows us to partially differentiate motor activity which is related with interaction patterns from that which results from user locomotion or other external variables.

- **Interaction Logging:** collects relevant data of user interaction with the application. The module encompasses metrics which may range from number of clicks on a specified interval, idle time to number of hops to reach a certain feature. Coupling these metrics with physiological data is paramount to identify emotional or motor fluctuations which are a result of interacting with applications or devices, from those which stem from external interventions.

- **Adaptation Core:** composed by data structures which store gathered data and a set of indexes and weights for each source, indicating how influential the data source is to the application or a control.
- **Event Manager:** provides a set of services which ensure the separation between all engine logic from the user interface, by using the Windows Message Queue as a channel.

3 Envisioned Scenario

The following example is a mere possibility of applying the engine in domains with widespread popularity. Entertainment, and in particular videogames, is a well suited thematic for adaptation examples.

3.1 Motion Game

Let us imagine a game in which players need to perform a series of gestures in order to accomplish certain tasks – a common setting with the increased popularity of motion-based interaction controls from major manufacturers. In this game, John, needs to perform the gesture in a fixed time window (see Figure 2) to cast a spell and defeat an enemy.

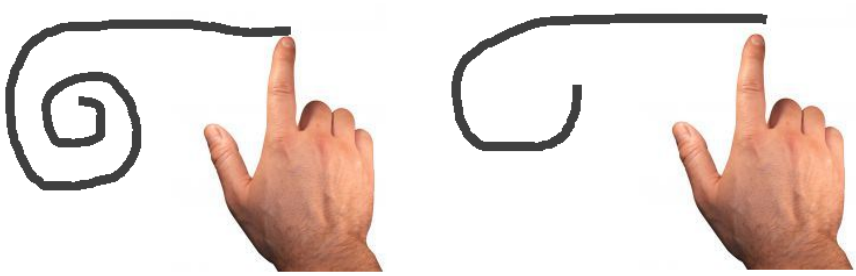


Fig. 2. Adaptation Engine's architecture

However, John is feeling tired from work and is currently struggling to achieve the desired motion in the appropriate time frame. While his accuracy is being recorded, the adaptation engine was capable of checking that his heartbeat rate has significantly increased since he started attempting to perform the gesture. Given the overall setting of a high biological signal being verified and diminished user accuracy, the game adapts itself to allow for simple gestures to be inputted.

This is a simple example of an adaptation process which encompasses user physiological data and interaction logging to change the way an individual interacts. Other possible scenarios may include difficulty adaptation, automatic scaling of interaction controls, interactive alternatives to the primary input methods, etc.

4 Future Work

It is our belief that as, a few years ago, accelerometers and other sensors were integrated into our everyday lives' devices, so will physiological interaction mechanisms in the future. The contribution from this ongoing work is far from accurately identifying emotions or exploring how to integrate physiological sensors into cell phones, but rather exploring scenarios and understanding user behavior and interaction patterns so that application design is tailored to our needs.

Future work encompasses finalizing the implementation of the adaptation engine and carrying our various experiments to assess how users interact in order to provide more realistic and tailored adaptation procedures in the engine.

Acknowledgments. This work was funded by FCT, through Individual Scholarship SFRH / BD / 39496 / 2007, through project PTDC/EIA-EIA/103676/2008 (InSIThe) and the Multiannual Funding Programme.

References

1. Duarte, L., et al.: Physiological Data Gathering in Mobile Environments. In: Procs. of UBICOMP 2010, Copenhagen, Denmark (2010)
2. Kwon, D., et al.: Combining body sensors and visual sensors for motion training. In: Procs. of ACE 2005, pp. 94–101. ACM, New York (2005)
3. Meehan, M., et al.: Physiological measures of presence in stressful virtual environments. In: Procs. of SIGGRAPH 2002, pp. 645–652. ACM, New York (2002)
4. NikeRunning, http://nikerunning.nike.com
5. Rowe, D., et al.: Heart rate variability: indicator of user state as an aid for human-computer interaction. In: Procs. of CHI 1998, pp. 480–487. ACM, New York (1998)
6. Shi, Y., et al.: Galvanic skin response (GSR) as an index of cognitive load. In: CHI EA 2007, pp. 2651–2656. ACM, New York (2007)
7. Tomida, T., et al.: The communication entertainment content by using entrainment phenomenon and bio-feedback. In: Procs. of ACE 2007, pp. 286–287. ACM, New York (2007)
8. Vaara, E., et al.: Mirroring bodily experiences over time. In: CHI EA 2009, pp. 4471–4476. ACM, New York (2009)

Dual Flow Interaction: Scene Flow and Data Flow, Dual Interaction in Art Installations

José Mª Alonso-Calero[1], Arcadio Reyes-Lecuona[2],
Jesus Marín-Clavijo[1], and Josefa Cano-García[1]

[1] Faculty of Fine Arts, University of Málaga
[2] Department of Electronic Technology, University of Málaga

Abstract. In an interdisciplinary context with regards to the experience of interaction in virtual reality art installations, we propose an analysis from the point of view of exchanges between the user and the art installation, which is produced by a dual-flow: data flow and scene flow. This dual flow is defined as a double layer which in parallel has a physical component and an abstract one, the division is between a physical interaction considered as external and the virtual as internal. The purpose is to identify and focus the relationship between materials and components of the interaction experience.

Keywords: Art installation , virtual art, interaction, physical interface, virtual reality, space, scenic, scene flow, participation, spectator-user presence.

1 Introduction

This communication is part of an interdisciplinary research effort involving work between artists and engineers on fields and areas we have in common. That is why we tackle the issue of materials and terminology, which may very well be defined in engineering and have a less prosaic use in the arts. In the field of art installation, the incorporation of technology has led to interest in participation and interaction, and has produced, among others, new concepts such as open book or dematerialization of art. In this paper, we show a diagram of the proposed relationship between these two areas, with different objectives and language, to reorganize aspects and components of the interaction experience.

1.1 Duality User-Spectator

The evolution of media and artistic expression has been responding to a natural demand by the spectator, who engages in the artistic event. His demand is implemented as an intervention in the production with the intention of being part of a work of art through a meaningful event leaving open the work of art in the process. He wants to interact and exchange his role with the artist.

1.2 From Representation to Participation

In the field of contemporary art, conceptual art in its reductionism, leads to the dematerialization of art where the object is voided for the live experience where

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 503–506, 2011.

space and time take center stage. The artistic practices of conceptual art pioneers reflected proposals with a clear attitude towards the environment and audience participation [1].

Frank Popper links the term environment to the intervention of the spectator, where time determines the spectator's participation [2].Participation as a possibility of intervention in order to alter, manipulate and transform everything. This experience is progressively built as a result of our actions in the environment. In the words of Piaget *intelligence should not contemplate but instead, transform* [3].From the point of view of constructivist psychology *the world of experience is always, exclusively a world constructed with the concepts we produce* [4]. For Peter Weibel the interactive nature of media art consists of three elements [5]: virtuality, variability and viability which correspond to how information is stored, user modification, and how everything is modified according to image behavior.

1.3 Contextualization

In this paper, we focus on the interaction that occurs in art installations with Virtual Reality, but understood in a broad sense [6]. We do not want to focus on the concept of the classic 3D Virtual Reality environment with which it interacts by browsing, selecting and manipulating objects [7]. We understand that this model is applicable to any art installation where, in one way or another, the spectator interacts with a physical space that has an abstract equivalent.

During the research of contrasting empirical aspects in these art installation forms we rethink the issues related to participation in the work of art and presence in the interaction.

2 The Dual-Flow Interaction

In the process of interaction, in addition to the physical aspect there is an element of abstraction, which is symbolic and semiotic, related to the information the user carries, what we would call cultural baggage similar to that mentioned by David Nuñez in the sense of presence in a virtual environment. David referred to it when he speaks of the user, to the extent that when the user feels he is in the virtual environment he is influenced by two things: first what comes to him through the physical stimuli and second, the life experiences he already has, his cultural baggage [8].

These two elements are present throughout the whole course of the interaction experience, through what we consider Dual Flow is composed of:

1. Data flow, which is information and data at the physical level.
2. Scene flow which moves at the abstract level.

2.1 Scene Flow

In the interaction experience, we believe that on the superfluous level there is a load of abstract content which we lay as a superimposed layer. In this layer of abstraction, aspects of language art and semiotic language are referenced. We defined scene flow as the catalyst between the user and the physical installation, which is presented through this layer of abstraction, on a psychological level. Scene flow would be the communication of abstract information.

We assume the fact that the communication between layers is straightforward, it is not necessary to be aware of the process of encoding, transmitting and decoding. It is a virtual communication, unlike data flow, that provides neither encryption or decryption, or physical transmission as direct communication, such as telepathy, between abstraction layer and the upper level which houses the physical installation and the user.

A further analysis of the same idea is done by E. Couchot, who describes the difference between internal and external interaction, external to the human-machine interaction and internal to the relationship with the virtual world elements [9]. Depending on the place of reference for the user, we can define two kinds of interaction, an internal (virtual world) and external (physical world).

Virtual reality has always sought to promote the internal interaction and minimize the external. Its classic paradigm is based on tricking the user to perceive the virtual environment as the real world. Minimizing the user's perception that this is an external interaction, its goal is to ensure that he does not realize he is interacting with devices. With that objective, a dramatic protocol is created inviting spectators on a journey to engage them in the experience [10]. In this process we must incorporate metaphors and empirical models to the artistic approach, where we discuss the use of empirical methods in the field of art.

2.2 Data Flow

Data flow operates in a similar way to the physical level of network protocol towers, sending packaged information that is encoded and decoded from one side to another in the process. Such codification would be the process of communication between different layers. Once coded and divided into smaller packages, information passes from one layer to another. When this information reaches the destination, it is decoded again to reach the next level.

3 The Layers of Abstraction

In this paper we proposed a diagram showing the location of the concepts and materials relating to the interaction. In our diagram, art and technology are at the ends, one on each side, separated by the interface which in turn divides the experience between physical interaction (external) and virtual interaction (internal), these two interactions are of two spaces : physical and virtual. Dual flow would function in the opposite way as the interface, the interface as its name says is between two sides but the flow is that which flows from one side to the other of the interface, is what flows from one space to another.

The abstraction layer is a superimposed structure where communication is on a more abstract level, speaks more of art, psychology concepts that are at a higher abstract.

4 Conclusions

In summary, in this paper we intend based on the interdisciplinary subjects shared by both artists and engineers, to identify the transfer and exchange relations present in the experience of interaction and physical interaction, which are combined as follows: user – installation, physical interaction – interaction virtual, real space – virtual environment, abstract information – information data, data flow – scene flow.

Our proposal is presented through the relationship diagrams where dual flow is between the installation and the spectator, data flow moves at the reality level and has a physical and technological nature, whereas scene flow moves at the virtual level and is abstract in nature.

References

1. Popper, F.: Arte, acción y participación: El artista y la creatividad de hoy. In: Akal / Arte y Estética (ed.), Madrid, p. 21 (1989)
2. Popper, F.: Arte, acción y participación: El artista y la creatividad de hoy. In: Akal / Arte y Estética (ed.), Madrid, p. 37 (1989)
3. Piaget, J.: Psicología y epistemología. Planeta-De Agostini, S.A., Barcelona, p. 89 (1985)
4. Von-Glasersfeld, E.: Despedida de la objetividad en El ojo del observador. Contribuciones al constructivismo. Editorial Gedisa, S.A., Barcelona, p. 27 (1994)
5. Weibel, P.: La irrazonable efectividad de la convergencia metodológica del arte y la ciencia. Epifanía, Marce-lí Antúnez Roca, coord. Claudia Giannetti. In: Arte facto y ciencia (ed.) Fundación Telefonica, Madrid, p. 54 (1999)
6. Timón, E.: Los problemas epistemológicos de la realidad virtual. Congreso Internacional de Filosofía y Realidad Virtual, celebrado en Albarracín. Teruel (2002)
7. Bowman, D.A., Kruijff, E., LaViola, J.J., Poupyrev, I.: 3D User Interfaces: Theory and Practice. Addison-Wesley Professional, Boston (2004)
8. Nunez, D.: A constructionist cognitive model of virtual presence. Collaborative Visual Computing Laboratory. CS04-07-00. Department of Computer Science. University of Cape Town, August 18 (2004)
9. Gianetti, C.: Estética digital: Sintopía del arte, la ciencia y la tecnología. L'angelot. Giuliano, Gustavo, Barcelona, p .119 (2007)
10. Lorenzo, R.: Filosofia y realidad virtual. In: Moreno, C., Lorenzo, R., De-Mingo, A. (eds.) Instituto de Estudios Turolenses. Prensas Universitarias de Zaragoza, Zaragoza (2007)

Effects of Touch Screen Response Time on Psychological State and Task Performance

Nozomi Sato[1] and Kentaro Nakajima[2]

[1] Kinki University, Faculty of Applied Sociology,
3-4-1 Kowakae, Higashi-Osaka, Osaka 577-8502 Japan
[2] Fuji Xerox Co., Ltd, 9-7-3, Akasaka Minato-ku, Tokyo 107-0052 Japan
nozomi.satoh@socio.kindai.ac.jp

Abstract. The purpose of this study was to investigate how touch screen response time affected user's psychological state and task performance. Ten male participants performed numeric entry task on the touch screen under six different conditions by crossing speed of touch screen response for pressing buttons and of switching pages. Results suggested that the touch screen with the faster response time would be accepted more favorably than that with the slower response time. However, with regard to the results of task performance, opposite trend was obtained.

Keywords: touch screen response, psychological state, task performance.

1 Introduction

Touch screen technology has become highly popular in our everyday life. It has been adopted not only for personal use such as home computer and audio equipment, but also for public use such as vending machine and automated teller machine. To improve user's psychological comfort in touch screen use, appropriate response time of touch screen for user's input is one of the essential issues. Time delay is a major source of annoyance with computing [1]. Much technological effort has been concentrated on shortening response time with computing. However, few studies have been conducted to investigate the negative effects of short response time of tough screen on user's psychological states and task performance. Therefore, the purpose of this study was to examine how touch screen response for pressing buttons and switching pages affected user's psychological state and task performance.

2 Methods

2.1 Participants

Ten male undergraduate students aged from 21to 24 years participated in this study. All participants were right-handed and had previous touch screen experience. They had normal or corrected to normal visual acuity and had no history of musculoskeletal diseases. Before enrolling in this study, written informed consent was obtained from all participants.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 507–510, 2011.

2.2 Experimental Task

Participants performed the numeric entry task on the touch screen, using the forefinger (Fig. 1). The task was a 10-digit input with numerical keypad on the 17-inch LCD touch screen (NT-7106, Nippon tect, Ltd.). 10-digits were displayed in the field on the top of the screen (Fig. 2). Participants were required to input the displayed digits with the numerical keypad. Actual digits that participants entered were displayed in the field in the middle of the screen (Fig. 3). Each trial was completed by pressing the button of the field (dotted line; Fig. 3). Six conditions were created by crossing speed of touch screen response for pressing buttons (1sec and 2sec), and speed of switching pages (1sec, 5sec, and 10sec). Participants completed whole conditions, each consisting of 10 trials.

2.3 Measures

Performance: To evaluate task performance, entry time for 10-digits of each trial was calculated by reducing touch screen response time from the time to complete the trial. Correct rate was also calculated for each condition.

Subjective time estimation: Previous study suggested that errors in estimation of durations could indicate the mental workload of the task [2]. Therefore, to evaluate mental workload while manipulating touch screen, subjective time estimation was conducted. Before the first condition and after each condition, participants were required to press a button of the stopwatch for a time period subjectively equal to 10 seconds.

Fig. 1. Participant conducting entry task

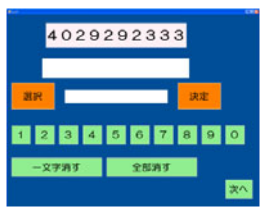


Fig. 2. Layout of the touch screen

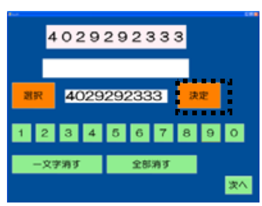


Fig. 3. An example of entered digits on the screen

Behavioral observation: Behavioral observation is one of the useful methods for evaluating usability [3]. Participants' behavior while the task was recorded by the digital video camera (PCR-TRV70k, Sony Corporation) to investigate participants' psychological state such as irritation and hesitation.

Questionnaire: To evaluate participants' preference for each condition, SD method (semantic differential method) was conducted with 14 items, which was used in the previous study [4]. Responses were made on a 7-point scale from 1(strongly non-preferable) to 7(strongly preferable).

2.4 Experimental Procedure

Following a brief training session, participants conducted subjective time estimation as a baseline. Then, they completed the numeric entry task under the six conditions. At the end of each of condition, they conducted the subjective time estimation and filled out the questionnaire. The order of condition was randomized for each participant.

3 Results

Performance: Entry time under different speed conditions for touch screen response time and the switching pages is presented in the Fig. 4. There was a significant main effect for speed of touch screen response ($F = 42.19$ [1, 9], $p<0.01$). The entry time for 2sec touch screen was faster than that for 1sec touch screen. No significant main effect and interaction was found for correct rate.

Subjective time estimation: Subjective time estimation under different speed conditions of touch screen response time is presented in the Fig. 5. There was a significant main effect for speed of touch screen response ($F = 8.90$ [1, 9], $p<0.05$). Participants estimated the time longer after the condition of 2sec response time than that of 1sec response time.

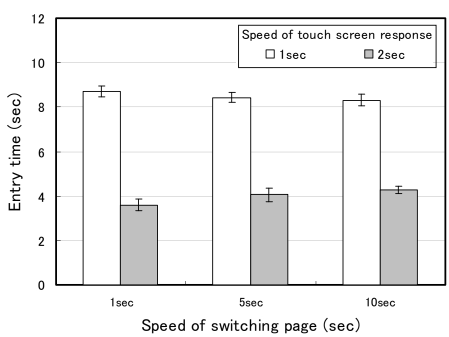


Fig. 4. Entry time for each condition. The error bars represent ± standard error of the mean.

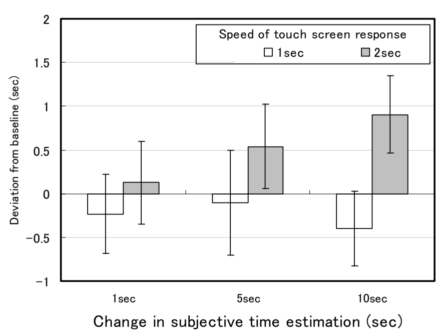


Fig. 5. Subjective time estimation after each condition. The error bars represent ± standard error of the mean.

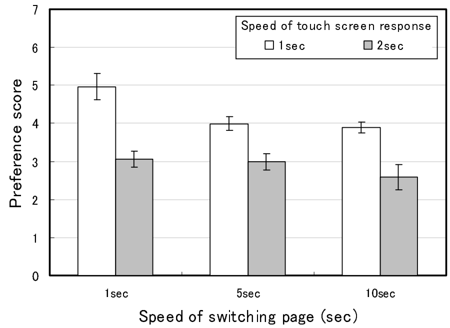


Fig. 6. Preference score for each condition. The error bars represent ± standard error of the mean.

Questionnaire: Preference score was obtained by averaging the total score of 14 items. Fig. 6 shows the preference score for each condition. Higher scores indicate greater preference. Friedman test was conducted to analyze data. There was a significant differences in preferred speed of touch screen response time and switching pages. Participants preferred touch screen response time with 1sec to that with 2sec ($p<0.01$), and preferred switching page speed with 1sec to other two speed conditions ($p<0.05$).

Behavioral observation: Statistically, there were no significant differences in the conditions. However, it was observed that some participants sighed or expressed restless behaviors such as tapping toe and shaking upper body slightly while response time and/or switching pages of touch screen were slow.

4 Discussions

Results from time estimation and questionnaire suggested that the time delay of touch screen button may negatively affect user's psychological state. Participants could have accepted more favorably the touch screen which responds in a shorter duration. However, task performance of the touch screen which responds with 1sec was worse than that with 2sec. Under the faster response time condition, participant had no enough time to move their fingers to the location of the button to be pressed although the system had been ready for accepting the entry. Changing of the layout of numerical keypad and/or adjusting the sensitivity of the touch screen would solve this problem. In the further study, the number of participants should be increased to confirm the reliability of results obtained from the current study.

References

1. Branaghan, R.J., Sanchez, C.A.: Feedback preference and impressions of waiting. Human Factors 51, 528–538 (2009)
2. Lind, M., Sundvall, H.: Time estimation as a measure of mental workload. In: Harris, D. (ed.) HCII 2007 and EPCE 2007. LNCS (LNAI), vol. 4562, pp. 359–365. Springer, Heidelberg (2007)
3. Huang, H., Lai, H.-H.: Factors influencing the usability of icons in the LCD touchscreen. Displays 29, 339–344 (2008)
4. Komatsubara, A., Yokomizo, Y., Yamamoto, S., Noro, K.: Analysis of work load by computer system response time in different type of VDT tasks. The Japanese Journal of Ergonomics 21, 211–217 (1985)

Elaborating Analysis Models with Tool Support

Gregor Buchholz and Peter Forbrig

University of Rostock,
Institute of Computer Science,
Albert Einstein Str. 21,
18059 Rostock, Germany
{gregor.buchholz,peter.forbrig}@uni-rostock.de

Abstract. Integrating models as essential elements into the software development process is supported by numerous methods and tools but the creation of such models still bears a considerable challenge. This paper proposes a structured modeling of tasks and activities during the requirements analysis in order to pave the way for the very early utilization of models. A tool implementation demonstrates the elaboration of models based on scenarios.

Keywords: Task Models, Scenario Analysis, Model Based Development.

1 Introduction

1.1 Motivation

A considerable number of methods and tools for the integration of task models into software development and usability testing can be found. The underlying paradigms (model-based and model-driven development) include the use of models in all development stages [4]. Models for design have to be deduced from requirements analysis results which would be much more convenient if the analysis stage itself purposed the formalization of its output as a model. While it might be a feasible objective to model user intentions, goals and tasks at a high level it becomes more and more difficult and laborious as the level of detail grows [7]. We will introduce a tool supported method to create such models during the requirement analysis phase.

1.2 Related Work

Deriving descriptions of user and system tasks from recorded scenarios has been conducted and documented before. The *CelLEST* process [9] reconstructs a navigation model of an existing application as state graph with screenshots as states and user inputs as state transitions, aiming at the documentation of existing systems. In [5] *ActionStreams* is proposed for the creation of task models based on user behavior. The users' activities are observed and recorded for a comparatively long time ("a little more than a day") and a grammar production represents the autonomous learning of the system. An example for the analysis of interaction sequences to deduce domain knowledge for learning systems via *Sequential Pattern Mining* (SPM) is the framework described in [3] that has been proved in the tutoring system *RomanTutor*.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 511–514, 2011.

The *ProM* framework [2] features a wide variety of algorithms to discover and mine processes and a lot more.

In addition to the construction of a formal correct model our approach includes the semantic meaningful abstractions of actions to task.

2 Collaborative Model Building as Learning Activity

The goal of the analysis phase is to define the problem to be solved and the desirable outcome is a clear and precise description of tasks for which software support has to be built. That elaboration of a common understanding of the problem is challenging and methods and artifacts are needed that help to arrange the acquired pieces of knowledge, to reveal missing information and to support discussion.

2.1 Gathering Domain Knowledge

This work supports the idea of the employment of task models as domain-independent modeling technique during all phases of software development. The selection of techniques for the first requirements elicitation is rather unaffected by the intention of task modeling. As usual, functional as well as non-functional requirements are wanted. That includes the identification of user roles, business objects, context information, use cases etc. and also descriptions of scenarios.

2.2 Structuring Domain Knowledge

The formal specification is started with the main goal of the problem to be solved on an abstract level which is then refined into more concrete tasks. Temporal relations between tasks can be modeled with the operators known from CTT (*Concurrent Task Trees*) [8]; missing operators will be deduced later. Each model can be assigned to a role as a separation into distinct models for different roles may be useful.

The requirements elicitation not only unveils abstract tasks but will also comprise more concrete actions. These actions can be assigned as sub-elements to tasks, so as to build groups and specify which section an action belongs to.

2.3 Scenario-Based Exemplifying

In the next step the models with tasks and actions are used to generate examples of scenarios. The developed tool supports the role-based recording of actions as well as a file import functionality to include traces produced by other systems, such as a log file from an existing software system.

2.4 Construction of Detailed Models

A tree with abstract tasks ("Manage account", and "Use account" in Fig. 1), a set of actions assigned to these tasks, and a set of action traces exemplifying the fulfillment of the tasks (see Fig. 1) are the basis for model generation. The resulting tree structure is a candidate for suitable representation of processes the system is to support (see the tree in Fig. 1). The tool features three steps of model generation:

Pattern recognition identifies frequently occurring sequences of actions in the traces based on the assumption, that actions observed together may be steps of a more abstract task. For finding such patterns the *Apriori* algorithm, *PrefixSpan*, and *BIDE+* are implemented. The recognition results in groups of potentially related actions. A reviewing of these groups comprises a reasonableness check and labeling.

The goal of **hierarchy creation** is to combine the manually *top-down* decomposed tasks with the generated parts created *bottom-up*, only loosely coupled to the upper part. In short, this step merges analytic and synthetic parts using a *LearnModel* algorithm modification without objects assigned to tasks and with the notation and operators of CTT until all actions from the protocols are assigned directly (one edge connecting both nodes) or indirectly (one or more nodes between the nodes) to a task.

In the process of **identifying temporal relations** initially all sibling nodes are considered to be in an *enabling* relation following their first occurrence in the first trace of the analyzed trace set they occur respectively. The protocol analysis corrects this assumption and *order independence, iterations* and *concurrency* are suggested. Each time an operator is modified all operators up to the root are modified, if violated.

This process can be repeated any number of times, selecting different sets of task traces and with varying settings for operator detection. Thus, different rival model versions can be created and discussed.

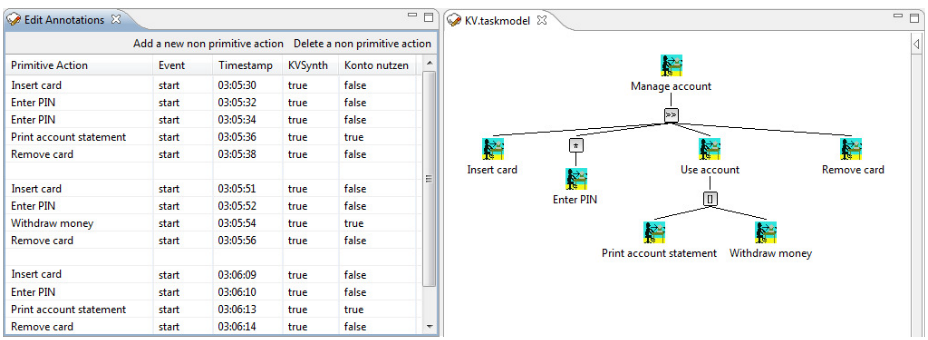


Fig. 1. Sample scenario traces and resulting model

2.5 Evaluating the Progress and Starting of New Iteration

The models can be simulated by the provided tool. Therefore, one or more clients connect to the server; a stored model is selected and opened in the simulation module with a specific role. Then, the execution of actions is simulated by activating the actions in the wanted order according to the previously detected scenarios. If any needed action is not accessible (that is, it is not in the enabled task set) or an action is enabled but known to be forbidden at that time, the model can be corrected. These simulations' protocols can serve as input for the next model generation iteration.

3 Conclusion and Future Work

This paper describes a method to establish models as an integral part of requirements analysis. Traces of scenarios can be combined with the structured knowledge about tasks to formalize the findings about the current situation. The approach employs modified data mining and association algorithms to derive relations between actions and tasks and build a hierarchical task structure. Results of the process have to be interpreted as a model of an existing situation. The approach and the tool support may help in simplifying the transfer of previously informal knowledge to structured and formal presentations.

Next steps will be the improvement of tool for annotations and references to other artifacts. Furthermore, it has to be investigated in which way the implemented tool can contribute to test prospected scenarios. In this situation, not the generation but the validation of models will be in focus.

References

1. Colombo, P., Khendek, F., Lavazza, L.: Requirements Analysis and Modeling with Problem Frames and SysML: A Case Study. In: Kühne, T., Selic, B., Gervais, M.-P., Terrier, F. (eds.) ECMFA 2010. LNCS, vol. 6138, pp. 74–89. Springer, Heidelberg (2010)
2. van Dongen, B., Alves de Medeiros, H.M.W., Verbeek, A.J.M.M., van der Aalst Weijters, W.M.P.: The ProM framework: A New Era in Process Mining Tool Support. In: Ciardo, G., Darondeau, P. (eds.) ICATPN 2005. LNCS, vol. 3536, pp. 444–454. Springer, Heidelberg (2005)
3. Fournier-Viger, P., Nkambou, R., Nguifo, E.M.: A Knowledge Discovery Framework for Learning Task Models from User Interactions in Intelligent Tutoring Systems. In: Gelbukh, A., Reyes-Garcia, C.A. (eds.) MICAI 2006. LNCS (LNAI), vol. 4293, pp. 765–778. Springer, Heidelberg (2006)
4. Johnson, S.K., Brown, A.W.: A Model-Driven Development Approach to Creating Service-Oriented Solutions. In: Dan, A., Lamersdorf, W. (eds.) ICSOC 2006. LNCS, vol. 4294, pp. 624–636. Springer, Heidelberg (2006)
5. Maulsby, D.: Inductive Task Modeling for User Interface Customization. In: Proc. IUI 1997, pp. 233–236. ACM, New York (1997)
6. Mazón, J.-N., Trujillo, J.: A Model-Driven Goal-Oriented Requirement Engineering Approach for Data Warehouses. In: Advances in Conceptual Modeling & Foundations and Applications, pp. 255–264. Springer, Berlin (2007)
7. Paris, C., Lu, S., Linden, K.V.: Environments for the Construction and Use of Task Models. In: The Handbook of Task Analysis for Human-Computer Interaction, pp. 467–482. Lawrence Erlbaum Associates, Mahwah (2004)
8. Paternò, F.: Model-Based Design and Evaluation of Interactive Applications. Springer, Heidelberg (1999)
9. El-Ramly, M., Stroulia, E., Sorenson, P.: Recovering Software Requirements from System-user Interaction Traces. In: Proc. SEKE 2002, pp. 447–454. ACM Press, New York (2002)
10. Seyff, N.: Exploring how to use scenarios to discover requirements. In: Requirements Engineering, pp. 91–111. Springer, London (2009)

End-User Support for Information Architecture Analysis in Interactive Web Applications

Luis A. Rojas and José A. Macías

Escuela Politécnica Superior, Universidad Autónoma de Madrid,
Tomás y Valiente 11, 28049 Madrid, España
luisalberto.rojas@estudiante.uam.es, j.macias@uam.es

Abstract. Information Architects analyze conceptual models and descriptions concerning non-functional requirements that will be later on used by Analysts and Software Engineers to design software artifacts. However, this flow of information is sometimes difficult to automate due to conceptual knowledge differences in the information processed by each expert. To facilitate this flow, in this paper we propose a CASE tool called InterArch. Our tool bridges the gap between both professionals, and it automatically generates UML diagrams for Analysts from the conceptual diagrams created by the Information Architect.

Keywords: Information Architecture, End-User Development, Analysis and Design of Interactive Web Systems.

1 Introduction

The Information Architecture (IA) is defined as the science of structuring, organizing and managing information, where labeling, findability and usability are specifically important for a good information performance [2]. Its main objective is to facilitate, as much as possible, the comprehension and assimilation of the information [1], as well as to execute tasks in user-defined information spaces [6].

The Information Architect attempts to find a bridge between the conceptual knowledge proposed by users in the problem domain and the design information that Software Engineers require to design the final interface in the solution domain. This gap frequently means that some degree of interoperability and alignment between the output generated by the Information Architect and the input expected by the Software Engineer is necessary in order to accomplish the whole software design. Furthermore, automating this flow of information could benefit from reducing both time and costs in software projects, also allowing each expert to focus on their own work and optimizing the transfer of knowledge among both professionals [5].

In order to address this problem, we have designed a CASE tool called InterArch, which allows experts in the problem domain to focus on content development and browsing. This way, our tool automatically generates UML content classes for Software Engineers from the conceptual models crated by Information Architects, thus automatically specifying the right elements for the solution domain.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 515–518, 2011.

2 Proposed Solution

Most common products that Information Architects have to create to carry out the IA of an interactive web applications correspond to *blueprints*, *wireframes*, *content models* and *controlled vocabularies* [2, 6]. From all this products, *content models* are particularly important for Analysts and Software Engineers, and they are very likely to be automatically processed in order to generate content classes and objects that define the solution in the application domain. This is the reason why we have focused on these essential products in favor of an explicit automation to provide useful input for the Software Engineer.

To address this challenge, our tool is based on two essential points: first, InterArch enables Information Architects to concentrate on conceptual-design tasks in the problem domain. Second, based on the initial analysis carried out by Information Architects, InterArch automatically generates UML diagrams for Analysts and Software Engineers, specifying elements that have a direct correspondence with the information content. UML information is generated in a transportable text-like format called XMI[1] that improve interoperability and can be processed by a great deal of existing CASE tools in order to accomplish the rest of the phases in the software project lifecycle.

2.1 A CASE Tool for the Information Architect

InterArch comprises a set of processes that are responsible for the management and transformation of information models in a visual environment for the IA analysis. These processes include: visual modeling of the conceptual elements required by the information professionals, transformation of the visual model into an intermediate model, and textual-transportable generation of the analysis of the information using UML diagrams. In a nutshell, these processes aim to take the input – i.e., the visual diagrams created by the Information Architect, and generate the output comprising UML diagrams for Analysts and Software Engineers.

The main idea behind these architectural components is to enable Information Architect to work in the visual processing of diagrams in a transparent way, but also featuring an interpretation overlay that recognizes the different correlations between such analysis diagrams and the UML classes required by Software Engineers. The transformation of the visual model is achieved through a set of relations and association rules which are applied to the conceptual model in order to generate the corresponding UML diagrams in XMI format.

Figure 1 depicts the user interface of InterArch, whose main parts are labeled with A, B and C. Part A includes a tool bar that implements functionalities concerning files, edition, format and style and others. Part B represents the main authoring environment used by the Information Architect. Finally, part C shows the icon library to enrich the visual composition of diagrams developed by the Information Architect.

[1] XMI is the OMG-defined standard for the exchange of UML diagrams, www.omg.org

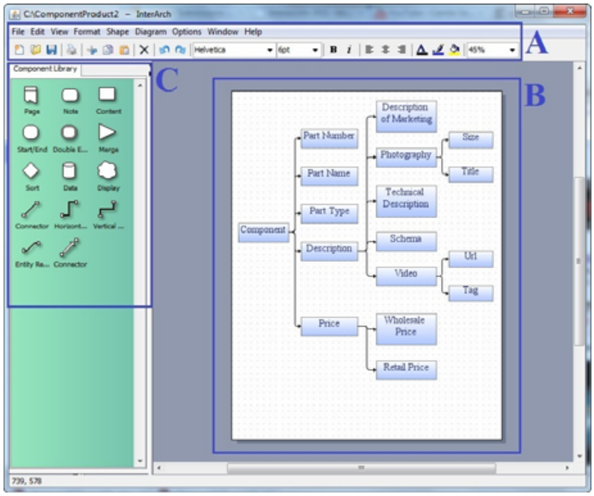


Fig. 1. InterArch User Interface

2.2 A Worked Example

In order to describe our tool in detail, we provide a specific example using InterArch. Let us suppose that the Information Architect is working on a content model including information about the composition and prices of products in order to carry out the IA analysis of an on-line shop. This situation corresponds to the diagram depicted in Figure 1.

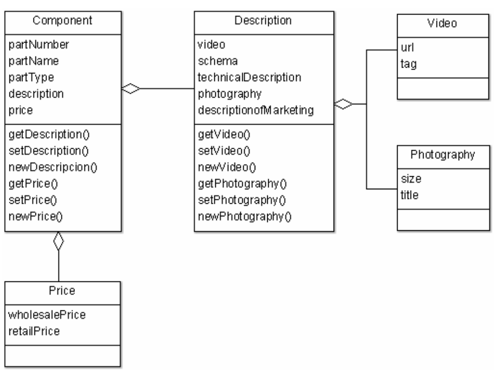


Fig. 2. UML Class Diagram generated by InterArch

To generate UML diagrams based on IA diagrams, transformation rules are applied in order to find out meaningful relationships. Carrying on with the example (see Figure 1), element **Component** will be first generated as a main class, also considering the descendant elements that will be processed by applying the following rule: terminal elements, and direct descendants from the main element, will become attributes for class **Component**. If these elements also have descendants, they will directly become new classes related to **Component**. This rule is recursively applied for all descendant elements in order to generate new classes, attributes and relationships. This way the application of this rule generates, for instance, a class **Component** with five attributes: **partNumber**, **partName**, **partType**, **description** and **price**. As for the class methods, three methods for each attribute that represents an aggregate class are generated by default (get, set and new). Figure 2 shows the final visualization of the UML classes generated including all the elements.

3 Conclusion and Future Work

In this paper, we propose an approach that provides an automatic solution to bridge the gap between high-level conceptual representations of content and the UML analysis and design classes needed to implement such content.

Our research is based on the EUD (End-User Development) paradigm [3, 5], where the main idea is to help end-users create or customize applications easily. More specifically, our work is based on EUSE (End-User Software Engineering), whose aim is to involve non-expert users, as often happens with Information Architects, in particular issues of the development process in software analysis and design [4].

As future work, we expect to provide new features, such as the generation and customization of rich design elements like design and content components. Also, we expect to carry out a usability evaluation, in order to study the trade-off between ease of use and expressiveness by involving professional Information Architects.

Acknowledgements. The work reported in this paper is being supported by the Spanish Ministry of Science and Technology, project ID: TIN2008-02081/TIN, the Madrid Research Council, project ID: S2009/TIC-1650 and, together with the UAM, project ID: CCG10-UAM/TIC-5772.

References

1. Elaine, G.T.: Information interaction: Providing a framework for information architecture. Journal of the American Society for Information Science and Technology 53(10), 855–862 (2002)
2. Erlin,Yunus, Y. A., Rahman, A.A.: The evolution of Information Architecture. In: International Symposium on Information Technology, ITSim 2008. , vol.4, pp.1–6 (2008)
3. Klann, M., Fit, F.: End-User Development. D1.1 Roadmap. In: Proceedings of the End User Development Workshop at CHI 2003 Conference, Ft. Lauderdale, Florida, USA, April 5-10 (2003)
4. Lieberman, H., Paternò, F., Wulf, V.: End-User Development. Human Computer Interaction Series. Springer, Heidelberg (2006)
5. Macías, J.A., Castells, P.: Providing end-user facilities to simplify ontology-driven web application authoring. In: Interacting with Computers, vol. 19(4), pp. 563–585. Elsevier, Amsterdam (2007)
6. Morville, P., Rosenfeld, L.: Information Architecture for the World Wide Web, 3rd edn. O'Reilly Media", O'Reilly Media Inc (2006)

Enriching Evaluation in Video Games

José L. González Sánchez[1], Rosa M. Gil Iranzo[2], and Francisco L. Gutiérrez Vela[1]

[1] GEDES, Universidad de Granada, Spain
{joseluisgs,fgutierr}@ugr.es
[2] GRIHO, Universitat de Lleida, Spain
rgil@griho.net

Abstract. One of the greatest challenges to the evaluation of UX in video games is to ascertain if the experience is appropriate for the game. Thus, it is necessary to know how to measure Playability in order to analyze, optimize and adapt it to the player's preferences. However, it is also important to remember that the final satisfaction of the user depends on his or her emotional response, social and cultural influences and hedonic properties of the interaction process with a video game. In this paper we present a way to perform a UX evaluation based on Playability by adding hedonic factors. The aim is to easily and cost-effectively analyze the UX in an agile video game development process.

Keywords: Playability, UX Evaluation, Emotions, Cross-Cultural Factors.

1 Introduction: Playability and UX in Videogames

Player eXperience (PX or User Experience in Video Games) depends not only on the product itself, but also on the user and the situation in which he or she uses the product [1]. Due to the nature and design of videogames, user experience is enriched by recreational, cultural, and other subjective factors that make analysis and evaluation difficult using traditional methods commonly used in interactive systems Playability is a term used in the design and analysis of video is a live topic in the scientific community; it has been studied from different points of view and with different objectives. A comprehensive source of the most important references about playability and user experience in video games research works can be found in [2,3]. In this work we use the Playability Model to characterize the PX in Video Games, which attempts to unify the playability characteristics mentioned in the previous references [4,5]. This model proposes the following attributes for defining the degree of PX: Satisfaction, Learnability, Effectiveness, Immersion, Motivation, Emotion and Socialization. Playability is defined as: *"a set of properties that describe the Player Experience using a specific game system whose main objective is to provide enjoyment and entertainment, by being credible and satisfying, when the player plays alone or in company"*, Also, Playability represents *the degree to which specified users can achieve specified goals with effectiveness, efficiency and, especially, satisfaction and fun in a playable context of use*.

The importance of the hedonic UX factor in video games justifies the need to enrich video game evaluation methods and game metrics with hedonic properties in order to analyze the degree of experience according to a social and cultural

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 519–522, 2011.

background. In this paper we use the concept of Playability [4] as a measure of PX with the objective of performing an analysis based on "rich" evaluation, guided by facets of Playability. The aim is to enhance the overall game experience by completing information about the PX and determining which video game elements have more influence on the final experience through the emotional response, playability factors and social and cultural background. To illustrate these objectives we present a practical example of how to evaluate the PX by obtaining pragmatic and hedonic information about it and determining which game elements have more influence on the overall experience. To achieve this, we present an evaluation methodology which allows us to easily and cost-effectively perform a quantitative /qualitative and pragmatic/hedonic Playability analysis to apply it in video game playable prototypes in order to improve the overall experience during an agile video game development process.

2 Practical Example of Rich Evaluation in Video Games

To illustrate how the UX in video games can be evaluated we performed a videogame analysis taking into consideration the following objectives: analyze quantitatively/qualitatively the degree of PX in an easy and cost-effective manner; test the effect of different video game elements on the overall PX and complete the functional assessment and objectives of PX with hedonic evaluations (emotional response or social/cultural influence). The evaluation of playability was carried out in a laboratory, in order to observe how people actually play a game. The evaluation was divided into four stages. *Pre-Test*: Questionnaires and a short test to obtain information about player profiles. These were completed with emotional information and multicultural background influences. *Test*: We collected information in real time about player experience while users played a video game. We used observation techniques to measure facial and corporal expressions or biometric constants. *Post-Test*: We gave players different questionnaires and interviews to complete with information, especially subjective information related to hedonic properties (hedonic properties are a crucial factor in motivation, emotion and satisfaction attributes). *Facets of Playability* guided these questionnaires. Reports: We obtained a number of reports about the PX with information about which playability attributes had more influence, or which type of elements were more valued by the players. We also obtained special emotional information and cultural information using Emocards (Emontional cards/emoticons). The game chosen as an example for evaluation was "Castlevania: Lords of Shadow" by MercurySteam and Konami. To obtain the maximum information about UX we used different player profiles: 'expert' (a person who is a good player, knows the game platform perfectly and is comfortable with difficult game challenges) and 'casual' (a person who plays infrequently and looks for quick entertainment). The experiment involved the participation of 35 student volunteers from different degrees courses students at University of Lleida, Spain.

From the Pre-Test, we were able collect information about the profile of participants using multimedia questionnaires (images, pictures, music, videos, etc.). The most significant results about *Pre-Test*: The majority of participants were male (75%) between 19 and 22 years old (90%). They were considered to be casual players (play approximately 5 hours per week, and have experience of only one games console or a mobile phone). They had knowledge of different gaming platforms, including a mobile and a desktop platform (90%). The preferences for different game genres were: adventure (≈60%) and fighting (≈30%). 87% preferred to play in company. To identify emotional state, we reproduced a fragment of game OST (Original Sound Tracks). For 83% the music inspired "action" and made the user feel "energetic". In analyzing the adequacy of the game hero and the context of the video game, 77% of students said that hero transmitted the feeling "honest", 43% reported that the most valuable characteristic was the eye expression and 15% said the face. These factors represent hedonic properties that produce better immersion in the video game and socio cultural background influences.

During the *Test*, users played the videogame. Their facial and body expression was recorded and their heart rate monitored [2,3,6]. The aim was to detect visceral and emotional reactions occurring during the interaction process. With the arrival of challenges at the climax of the level, stress increased, provoking surprise and agitation in the player (lifting of the eyebrows, ≈94%). The stress caused by the challenges increased the player's concentration. Players pressed the pad quicker and more violently (≈89%) than at the beginning of the game. At the climax of the level, two types of strategies were detected: a defense strategy, which was adopted by 100% of female users and 25% of men, who were intimidated by the game enemies and natural catastrophes. However, 75% of male users preferred a direct and violent confrontation. In both cases the degree of immersion was high, a factor, which was indicated by the increase in heart, beats per minute (≈18bpm at the climax of the level). Finally, stress became satisfaction, which was expressed as a slight smile or slight gasp (≈82%). It is interesting to note that the strategy depends on cultural factors such as gender and other factors, which may be related to the survival instinct of the user when faced with real challenges.

At the *Post-Test* we used informal interviews and questionnaires to obtain information about the player experience. The questions and heuristics are designed to extract information about the pragmatic and hedonic PX dimensions. This result was obtained thanks to the relation to the Playability Model, Facets and design of questionnaires [7]. One of the clearest results was that players were happiest with the game. This was reflected in the high values scored for the interactive facet, and ease of use of game controller (see learnability attribute). The positive result for 'casual' players meant a negative experience for 'expert' players, due to the excessive ease of play. Using the Emocards, we are able to analyze factors that are difficult to assess objectively using a number on a scale of values. However, at the same time, players are able to easily identify the emotional impact they have felt. Post-Test results reaffirm the positive feedback from users with regards to the setting and context as well as sound effects and music in general (high level of satisfaction, mainly characterized by a state of pleasure, which varies between excitation and neutral state). On a global level, the UX results were very positive, and the game satisfied players. From the results, we consider that enriching the PX analysis with multicultural and emotional factors (hedonic properties) helps us to establish information about player preferences and likes. The techniques demonstrated in this work can assist in the improvement of a video game as a product by enhancing the final experience, where the cultural or emotional impact can influence how a game should be designed or modified according to different player profiles or the market in a cost-effective manner.

3 Conclusion

This work reflects the importance of analyzing the experience that players have with video games in a pragmatic and hedonic way. Understanding and evaluating the user experience in video games is important for developing more efficient and successful products in terms of entertainment where cultural influence and emotional impact are crucial properties. To acquire more information about the player experience, we propose the enrichment of tests with hedonic factors (techniques of emotional design and cultural analysis) to increase the efficiency of questionnaires and heuristics. An example of this includes the use of Emocards, tracking the player response in real time, and measuring biometric constants. The measurement and analysis process is considered to be easy and cost effective to perform, thus making it appropriate for agile development based on playable video game prototypes. Through a practical example, we have demonstrated the importance of this kind of evaluation for the development of a product that can incite better experiences or specific experiences depending on the target market and it is advisable to use in every interactive system. As a future work, user mental model will be incorporated in the evaluation process.

Acknowledgments. This research is financed by: the Spanish International Commission for Science and Technology (CICYT); the DESACO Project (TIN2008-06596-C02) and OMediaDis (TIN2008-06228). Thanks to MercurySteam for the support in this work.

Referentes

1. Law, E., et al.: Understanding, Scoping and Defining User Experience: A Survey Approach. In: Proc. Human Factors in Computing Systems (CHI 2009), pp. 719–728 (2009)
2. Isbister, K., Schaffer, N. (eds.): Game Usability: Advancing the Player Experience. Morgan Kaufmann, San Francisco (2008)
3. Bernhaupt, R. (ed.): Evaluating User Experience in Games: Concepts and Methods. Springer, Heidelberg (2010)
4. González Sánchez, J.L., Padilla Zea, N., Gutiérrez, F.L.: From Usability to Playability: Introduction to Player-Centred Video Game Development Process. In: Proc. Human-Computer Interaction International (HCII 2009). pp. 65–74 (2009)
5. González Sánchez, J.L., Padilla Zea, N., Gutiérrez, F.L.: Playability: How to Identify the Player Experience in a Video Game. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5726, pp. 356–359. Springer, Heidelberg (2009)
6. De Lera, E., Garreta-Domingo, M.: Ten Emotion Heuristics: Guidelines for Assessing the User's Affective Dimension Easily and Cost-Effectively. In: Lew, M., Sebe, N., Huang, T.S., Bakker, E.M. (eds.) HCI 2007. LNCS, vol. 4796, pp. 163–166. Springer, Heidelberg (2007)
7. González Sánchez, J.L.: Jugabilidad y Videojuegos. Análisis y Diseño de la Experiencia del Jugador en Sistemas Interactivos de Ocio Electrónico. Editorial Académica Española. Lambert Academic Publishing GmbH & Co. KG (2011)

Evaluation of 3D Object Manipulation on Multi-touch Surfaces Using Unconstrained Viewing Angles

Daniel Mendes and Alfredo Ferreira

INESC-ID / IST / Technical University of Lisbon,
Rua Alves Redol, 9,
1000-029 Lisboa, Portugal
{danielmendes,alfredo.ferreira}@ist.utl.pt

Abstract. Recently, considerable research has been carried out regarding three-dimensional object manipulation on multi-touch surfaces. However, most current solutions were developed having in mind scenarios with a camera perpendicular to a scene axis, and cannot be easily used to manipulate three-dimensional objects in unconstrained viewpoints. In this paper, we present and evaluate a set of object manipulation techniques. As a test bed for this study, we used an unconstrained virtual LEGO modeling tool, taking advantage of ongoing work and popularity of LEGO construction among people of all ages. From this evaluation we improved our understanding on how users prefer to manipulate 3D objects on multi-touch surfaces.

Keywords: Multi-touch interaction, 3D object manipulation, rotation, translation, user evaluation.

1 Introduction

From mobile devices to tabletops or wall-sized interaction surfaces, multi-touch is becoming part of our everyday life. Nowadays it is possible to build low-cost prototypes that can detect multiple simultaneous touches. This kind of technology is becoming more and more popular among the general public.

With the flourishing of these devices, new interaction paradigms are needed. Multi-touch interfaces can provide more natural and user-friendly interaction than the traditional WIMP interfaces. Indeed, studies show that multi-touch interfaces are also more efficient [6, 2]. Although new paradigms have already been defined in some scenarios, such as the manipulation of two-dimensional objects, other scenarios are in need of more work, such as the manipulation of three-dimensional objects.

Most research in this context were designed for scenes with a camera perpendicular to one of the axis, making it difficult to apply the proposed techniques in scenarios which allows the freely manipulation of the camera and need to precisely position and orientate an object. To have such scenario and address this challenge, we used a prototype of virtual LEGO models in a multi-touch table. In this paper, we propose a set of translation and rotation techniques, adapted from existing approaches, and evaluate their suitability for unconstrained viewpoints, comparing them with each other and with existing solutions.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 523–526, 2011.

2 Related Work

The research on object manipulation using multi-touch surfaces started with 2D objects. Hancock et al. [3] surveyed several techniques for both rotation and translation of this type of objects, using multi-touch based gestures. One of these techniques, the two-point rotation and translation, has become the *de facto* standard.

When manipulating of three-dimensional objects, the challenges are different. The main difference is the lack of direct mapping from input to visualization. Instead, it is necessary to map the 2D input to a 3D view. Several approaches were proposed to tackle this challenge. Hancock et al. [4] presented a study about manipulating 3D objects using one, two and three touches simultaneously. The approach using three touches showed the best results when tested and was later denoted as Sticky-Fingers [5]. Martinet et al. [7] developed two techniques to translate 3D objects: one extends the standard four viewports that can be found in many CAD applications, and another (Z-technique) uses only a perspective view of the scene.

The work presented above, regarding interaction with 3D objects on multi-touch surfaces, were made having in mind scenes where the camera does not move, or at least does not rotate, since the basic movement of the objects is done in a plane parallel to the viewing plane. Moreover, scenes with a different camera, such as an isometric perspective, will require additional cognitive effort from the user to understand the plane in which the objects are moving or, in other words, what is the relation between the viewing plane and the scene referential.

3 3D Object Manipulation on MT Surfaces

The test bed for our study was a prototype of virtual LEGO modeling on table-sized multi-touch surfaces, presented in [8]. It allows touch-based manipulation of blocks, rotating and moving them freely, to build virtual LEGO models and uses an orbiting camera. The techniques used for 3D object manipulation are derived from approaches presented above and are described next.

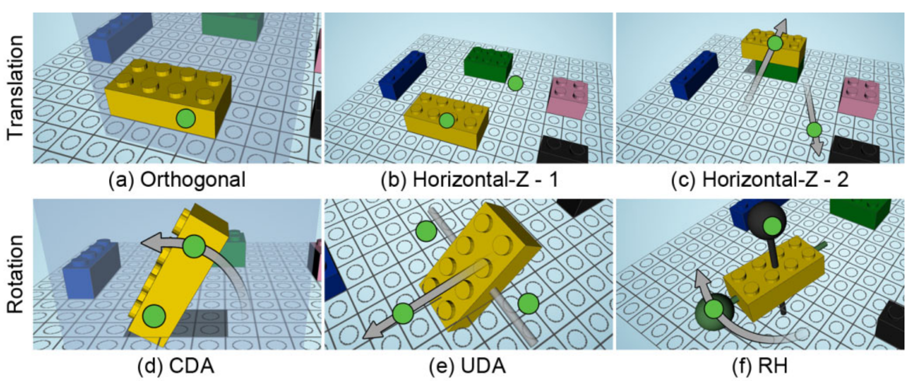


Fig. 1. Proposed 3D object manipulation techniques for unconstrained viewpoints. The green circles represent user touches and the gray arrows the movement of these touches.

3.1 Translation Techniques

Orthogonal – is based on the two-point rotation and translation for 2D objects [3]. To apply it to 3D, we use a plane orthogonal to one of the scene axes, closer to the current view vector to move the object (Fig.1.a). To use a different plane, the user will have to rotate the camera. We display a blue square representing the translation plane.

Horizontal-Z – derived from the Z-technique [7], but instead of moving the object in a plane parallel to the view, it moves the object in a horizontal plane. A second touch manipulates the depth of the object relatively to the camera (Fig.1.b,c). To improve the perception of object position, a shadow of the object being moved is cast.

Plane-Switch – uses a concept similar to the first technique. Initially, the translation is done on a horizontal plane. Touching on the blue square, the translation plane becomes the vertical plane that has the closer normal direction to the view vector. Thus, the user can move the object in 3D without changing the camera.

3.2 Rotation Techniques

Camera-Defined-Axis (CDA) – is used with the Orthogonal translation technique. While a touch is used to move the object, another touch is used to rotate the object (Fig.1.d). It uses the translation plane normal as rotation axis and the first touch as rotation center. To rotate around other axes, the user will need to rotate the camera.

User-Defined-Axis (UDA) – is similar to the Opposable Thumb [5]. The main differences are that the rotation axis is parallel to the X, Y or Z axis and the rotation center is the center of the object (Fig.1.e). The rotation axis used is the one closer to that defined by the two first touches. Thus, we can use the ease to define the rotation axis introduced by the Opposable Thumb and use a rotation axis easily discernible.

Rotation-Handles (RH) – uses the concept of virtual handles [1]. When the user selects an object, three handles parallel to the scene referential become visible. One touch selects the handle corresponding to the desired rotation axis, and another touch rotates the object around its center, by dragging other handle (Fig.1.f).

4 Comparative Evaluation

We set out to evaluate the techniques for 3D objects manipulation described above using a virtual LEGO modeling prototype. To verify our assumptions regarding translation using a plane parallel to the view, we included the Z-technique in our evaluation. There were two different tasks, one for translations, where user had to move five bricks until their position were coincident with ones with the same color but translucent, and another for rotations, where the user had to rotate four blocks to a specific orientation. Tests involved twenty users with very distinct backgrounds and a large (greater than A0) multi-touch enabled tabletop surface.

The times taken by each user to complete the tasks with the different techniques followed a normal distribution, accordingly to the Shapiro-Wilk test. We used the One-Way ANOVA test and the Post-hoc Tukey HSD multiple comparisons test to find out significant differences in the translation. The Z-Technique was significantly slower than the Orthogonal and the Plane-Switch. These two approaches were also faster than the Horizontal-Z, which had a much higher rate of failure to complete the task within the available time than all approaches. Concerning the rotation, we applied the T-test to each pair of techniques and concluded that the RH was significantly faster than the CDA and the UDA. The RH was also the one with the lowest rate of failure. We asked users to classify each technique in a four-point Likert scale and, by applying the Wilcoxon Signed Ranks Test, we observed that their preferences reflected what was concluded with task time and rate of failure analysis.

5 Conclusions and Future Work

We believe that existing solutions to 3D object manipulation on multi-touch surfaces work poorly with free viewing angles. In this paper we presented techniques for manipulating objects with this kind of scenario in mind. We tested these techniques with twenty users, comparing them among each other and with the Z-technique.

From this evaluation, we were able to conclude that users perform better with approaches that translate objects accordingly to planes perpendicular to one of the scene axes, instead of a plane parallel to the view. Even though a shadow of the object was casted to improve user perception of its location, some users initially thought that the gesture for depth manipulation was scaling the object, not moving, and found difficult to perceive the actual position of the object. We also observed that, when rotating objects, few users had some notions of what axis is needed to achieve a certain rotation. Most of them tried several axes until they identified the right one to use. We will address these issues in a near future.

Despite the evaluation with a LEGO modeling tool, we believe that conclusions here presented can be applied in generic 3D modeling scenarios. Our next step is to devise a set of manipulation techniques from the results obtained in this evaluation.

References

1. Conner, B., Snibbe, S., Herndon, K., Robbins, D., Zeleznik, R., Van Dam, A.: Three-dimensional widgets. In: Proc. of the 1992 Symposium on Interactive 3D Graphics (1992)
2. Fiorella, D., Sanna, A., Lamberti, F.: Multi-touch user interface evaluation for 3d object manipulation on mobile devices. Journal on Multimodal User Interfaces (2009)
3. Hancock, M.S., Carpendale, S., Vernier, F.D., Wigdor, D., Shen, C.: Rotation and translation mechanisms for tabletop interaction. In: Proc. Tabletop (2006)
4. Hancock, M., Carpendale, S., Cockburn, A.: Shallow-depth 3d interaction: design and evaluation of one-, two-and three-touch techniques. In: Proc. CHI (2007)
5. Hancock, M., ten Cate, T., Carpendale, S.: Sticky tools: Full 6dof force-based interaction for multi-touch tables. In: Proc. ITS (2009)
6. Kin, K., Agrawala, M., DeRose, T.: Determining the benefits of direct-touch, bimanual, and multifinger input on a multitouch workstation. In: Proc. GI 2009 (2009)
7. Martinet, A., Casiez, G., Grisoni, L.: The design and evaluation of 3d positioning techniques for multi-touch displays. In. In: IEEE Symposium on 3D User Interfaces (2010)
8. Mendes, D., Ferreira, A.: Virtual lego modelling on multi-touch surfaces. In: WSCG 2011 Full Papers Proceedings (2011)

Evaluation of an Accessible Home Control and Telecare System

Fausto Sainz[1], Judit Casacuberta[2], Marta Díaz[2], and Jaisiel Madrid[1]

[1] Fundosa Technosite S.A., Albazans, 16 3ª planta B1 28037 Madrid, España
[2] CETpD Technical Centre for Dependency Care and Autonomous Living, Barcelona Tech
Neàpolis Rbla. Exposició, 59-69 08800 Vilanova i la Geltrú, España
{fsainz,jmadrid}@technosite.es
{judit.casacuberta,marta.díaz}@upc.edu

Abstract. The article presents part of the research carried out within a project aimed at procuring interaction of people with disabilities and elderly with their environment through the use of information and communication technologies. We present methodological aspects related to participation models, user experience, technology acceptance and peer training. Technology was designed to test the effectiveness of systems and interfaces developed. Evaluation was conducted in an automation environment with older people as users.

Keywords: usability, accessibility, interactive system evaluation, evaluation methodology, device interaction, domotics.

1 Introduction

INREDIS [1] is a basic research project in the field of accessible and interoperable technologies aimed at developing basic technologies for communication and interaction between people with some kind of special needs and the environment. An architecture capable of adapting different types of interfaces to users' needs and preferences, was designed and developed. This paper explains how within the home environment, two use cases were assessed: a prototype home automation control and a telecare service. The home prototype allows the control of different home devices, while the telecare prototype offers to make an appointment request to different specialists.

1.1 Smart Home

A smart home is defined as a residence equipped with computer and information technology that meets occupants' needs, working to promote comfort, harmony, security and entertainment through the management of technology at home and its connections. Smart homes are also characterized by research focused on users [2]. Among essential factors considered to achieve a good experience are trust and perception of control. For trust to be relevant it should be two preconditions [3]: first there must be a dependency between the "truster" and "trustee", and this dependence must involve some risk, and further the risk must contain both uncertainty and

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 527–530, 2011.

vulnerability. According to [4] a comprehensive definition of trust comes from the union of three elements: the function or element that has to inspire confidence, the belief that anticipated results will occur and the will to act on this trust.

Conversely automation can be defined as the performance of functions by a machine agent (usually a computer), which have previously been carried out by an individual. Considering automation different ways of control were proposed [5] and applied in various technological areas: supervisory control, negotiated and shared.

2 Methodological Background and Evaluation method

In order to design a methodology adapted to the needs of the defined environmental assessment a study of two methodological approaches aimed assessing different products and services were conducted: Extended Technology Acceptance Model, and People Lead Innovation benchmark. The Technology Acceptance Model (TAM) [6] argues that attitudes towards the use of an information system is based on the perceived usefulness and perceived ease of use, it also suggests the link between perceived usefulness and intention to use in determining the use of the system. The Extended Model of Technology Acceptance [7] based on the original TAM and Innovation Diffusion Theory (IDT) [8], incorporates the dimensions of cost, and perceived risk. People Lead Innovation methodology [9] integrates users citizens in the innovation processes. The dimensions used in this methodology are: emotional approach, ergonomics, public innovation, sustainability and security management.

The validation of the automation environment involved evaluating the smart home control and monitoring system as well as the devices used to control the system. For each of the proposed elements a methodological adjustment was needed in order to determine what dimensions to assess during the experience. All platform elements - home control and telecare systems and devices- were evaluated and validated focusing on all those dimensions and material features that conform the usability and accessibility from a system. Also measures were taken related to information transmission, navigation, configuration, training, fitness requirements, ethics and privacy.

Information to assess the system was collected through post-test questionnaires (see Fig.1) measuring technology acceptance (comprise by perceived risk PR, cost COS, compatibility COM, perceived usefulness PU, perceived ease of use PEU and use intention UI), confidence CON (degree to which targeted users perceive the system as a reliable element for tasks' achievement) and automation level AUT(degree to which targeted users perceive that they control the system relative to automation level).

Participants. 12 elders aged between 70 and 88 years (mean = 79.33). Two had displacement problems and one had handling problems. 25% had previous experience with computers, and 18.18% with the Internet.

Devices. Four tactile devices: 1) iPhone, 2) iPad, 3) Tobii (used only as a touch screen computer), 4) touch screen PC. Each user was using a portable device and a non-portable device, always starting with the non-portable device.

Test Protocol. Firstly participants went into a devices familiarization and training period, followed by the test when they had to use the devices to perform the proposed tasks. Familiarization and training was carried out involving previous participants to observe knowledge transfer and identify this group representative vocabulary. The following tasks were performed: 1) INREDIS interface registration, it determines INREDIS profile type, 2) digital home use case (control of household items through the interface. e.g. up blinds, turn off lights, open door), 3) telecare use case (request, view and cancel a doctor's appointment).

Qualitative results. For iPad and touchscreen the most frequently encountered difficulties were "identification" and "navigation": identification of screen elements and functionality, menu navigation system, and how to move through the different screens. By activities Telecare services interacting with the iPad concentrated the largest number of use problems. For the iPhone and Tobii the difficulties encountered were concentrated in three activities: "Select items", "Identification" and "Accuracy needed". Regarding the difficulty of use identified by task, which can relate to the devices used, it could be seen that most problems were detected in the iPhone.

Quantitative results. Performance results indicated that all users were able to perform the tasks, although sometimes only partially: the telecare use case tasks were successfully overcome by all participants using the iPad.

Performance results with the iPhone and Tobii indicate that all users have exceeded the registering task without difficulty. However only one user managed using the iPhone to partially complete the telecare task (request, view and cancel a doctor's appointment). A significant amount of difficulties caused by the device's functions were detected, rather than system problems. The task of controlling the home elements was completed successfully by 2 users and partially by 5.

Questionnaires results. Likert scales1 to 5 were used to evaluate the devices and the system. In general the touch screen computer, with a mean of 2.09 obtained better scores than the iPad, with an average of 2.17. The dimensions most valued are the emotional aspects, with a mean of 1.77, followed by accessibility, with 1.97. The familiarity dimension was highlighted in relation to interaction, while cognitive ergonomics, which questioned navigation and information distribution obtained the worst rating, with an average of 2.34 and 2.33 respectively.

In relation to the system's perception in general it caused a good sensation and participants seemed to approve of it (see Table 1) considering that the system cover their daily needs at present time and would benefit from it in the future. They considered the system easy to use and with time they would use it proficiently.

Table 1. Results from the system questionnaire

CON	AUT	PU	PEU	UI	PR	COS	COM	Average
2.17	2.25	1.93	2.19	2.5	3.21	2.27	2.99	2.44

Peer training. Participants generally felt inhibited by having to explain how they used the devices, and according to their own utterances by the presence of the facilitators: "You explain it better" "I do not know, you better tell them". There was a tendency to explain the devices that they had been used more easily, mainly the Tobii as touch screen and the touch screen computer. The vocabulary used was not very specific e.g.: "You touch what you seek" "Tapping here?" referring press a button on the IPAD) and, in most cases, participants used gestures to indicate those elements of the system they needed to describe. Explanations lacked information regarding key aspects such as navigation and exploration of the interfaces.

3 Conclusions

There was a trend in participants not to assess negatively the devices, the applications neither the services on the usability questionnaires. Participants' comments showed they acted as if use was easy even though they needed to learn to use the system and devices: self-perceived as people with little technological knowledge. They tend to avoid criticism: very few users suggested changes to the system interface even when having major problems during its use. It would be interesting to think about a format that facilitates spontaneous expression of negative opinions and criticism.

Major problems were detected during the interaction with mobile devices iPhone and iPad, especially for features implemented in browsers' traditional interface, such as URL space, although the interface did not require using it, it puzzled users.

Cost of acquisition, installation, maintenance and support of the functionality offered by the system with users' real needs and perceived risk, specially the use of cameras in the home environment were detected as the main issues that could influence the acceptance and use of the system. In principle, factors that are not directly related to the interface or the system and unrelated to the interaction.

References

1. INREDIS, 2007-2010, `http://www.inredis.es` (accessed: March 30, 2011)
2. Van de Sluis, R., Eggen, J.H., Jansen, J., Kohar, H.: User Interface for an In-Home Environment. In: Hirose, M. (ed.) Human Computer Interaction - INTERACT 2001, pp. 383–390. IOS Press, Tokyo (2001)
3. Sandström, G.: Smart Homes and User Values: Long-term evaluation of IT-services in Residential and Single Family Dwellings (2009)
4. Chopra, K., Wallace, W.A.: Trust in Electronic Environments. In: Proceedings of the 36th Hawaii International Conference on System Sciences (HICSS 2003), January 6-9 (2003)
5. Sheridan, T.B., Parasuraman, R.: Human automation interaction. Reviews of Human Factors and Ergonomics 1, 89–129 (2006)
6. Davis, F.D.: Perceived usefulness, perceived ease of use, and user acceptance of information technologies. MIS Quarterly 13(3), 319–334 (1989)
7. Wu, J.-H., Wang, S.-C.: What drives mobile commerce? An Empirical Evaluation of the Revised Technology Acceptance Model
8. Rogers, E.M.: The Diffusion of Innovation, 5th edn. Free Press, New York (2003)
9. I2BC, Referencial People Led Innovation (2009)

Experimenting and Improving Perception of 3D Rotation-Based Transitions between 2D Visualizations

Maxime Cordeil[1], Christophe Hurter[1], and Stéphane Conversy[2]

[1] DSNA/DTI R&D, ENAC and IRIT/IHCS,
[2] ENAC and IRIT IHCS
7 Avenue Edouard Belin 31055 Toulouse
cordeil@cena.fr,
chirstophe.hurter@aviation-civile.gouv.fr,
stephane.conversy@enac.fr

Abstract. Exploring a multidimensional dataset with visualization requires to transition between points of view. In order to enable users to understand transitions, visualization can employ progressive 3D rotations. However, existing implementations of progressive 3D rotation exhibit some perception problems with visualization of cluttered scene. In this paper, we present a first experiment showing how existing 3D rotation is effective for tracking marks, and that cluttered scenes actually hinder perception of rotation. Then, we propose to set the axis of rotation on the graphical marks of interest, and ran a second experiment showing that focus-centered rotation improves perception of relative arrangement.

Keywords: Information Visualization, Visual exploration, Navigation, Transition, 3D Rotation.

1 Introduction

Air Traffic Control analysts have to analyze how a particular Air Traffic Controller proceeded to resolve a past conflict involving two aircraft. For these incidents analysis, they use two 2D-views presenting aircraft trajectories from top and vertical views. Thereby, they switch between these views in order to understand the resolution of the conflict, i.e. which aircraft has changed its flight level or heading.

The switch between views can use 3D rotation, a transition that supposedly helps track moving marks and their relative arrangement during the time of the transition. This type of animated transition consists of displaying a temporary 3D view that rotates like a transparent dice. It is used for example for exploration of scatterplots (ScatterDice [1]) or aircraft trajectories (FromDaDy [2]). The scatterdice-like transition implemented in visualization tools used by ATC analysts is slightly different; to explore data, users are free to pan and zoom 2D-views of trajectories. Since the rotation axis is always set at the X,Y,Z center of the data set, the scatterdice-like transition can provoke a loss of focus on these trajectories. Furthermore, in the original usage of scatterdice users deal with a small amount of data, whereas ATC analysts have to deal with a large amount of trajectories (up to 20,000), resulting in

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 531–534, 2011.

cluttered scenes. Previous studied benefits of animated transitions showed that rotation can help track objects [3] but the studies did not aim at assessing the perception of relative arrangement.

In this paper, we first present an experiment showing that 3D rotation can help users tracking objects and that cluttered scenes hinder perception of rotation. Then we propose an improvement of 3D rotation regarding problems we identified, and ran experiments showing that focus-centred rotation improves perception of relative spatial arrangement.

2 Experiment 1: Tracking Objects

The goal of this experiment is to validate that animated rotation helps users tracking graphical objects, and that density influences the effectiveness at this task. The underlying hypothesis is: (H1) higher density levels will negatively affect tracking objects.

The screen is a standard 21" LCD screen. A trajectory is depicted with a 1 pixel-width line. There were nine subjects, all regular computer users (researchers and PhDs in computer science, ergonomics, air traffic controllers), with an age ranging from 22 to 55 (average of age 40). Half of them are familiar with 3D environments. None are Infovis experts. Though the task is related to the specific activity, it does not require Air Traffic Control skills.

In this experiment, the system first presented a top view of green trajectories, with a particular red trajectory (the focus trajectory). After four seconds, the red trajectory turned to green and the system played a 3D-rotation from the top to the vertical view. In this last view, the same trajectory or another one was re-highlighted in red. Then, participants were asked to answer "Yes" or "No" to the question: "Is the red trajectory the same as the red one at the beginning of the animation?" During this first experiment, there were four levels of density (i.e. number of displayed trajectories): Empty (only the two trajectories of interest), Low (10 trajectories), Medium (20 trajectories) and Heavy (40 trajectories). The starting location of the trajectories is at the edge of the screen. We excluded the centre location since we assume that it helps the subject to track a mark. One of our concerns was that subjects would recognize trajectory configurations, and base their answers on previous trials. To avoid this effect, we varied the profiles of the trajectories, and the exact starting location. In summary, we tested 4 levels of density, with 4 trajectory profiles, at 4 locations. Thus, each subject performed 64 trials. The order of conditions was counter-balanced.

We merged the results according to density level and performed a measure-repeated ANOVA on the data. We only consider findings significant if the P value is less than .05. We first found that the level of density influences the correctness of the users' answers (p=,000). In the "empty" condition, users answered 88% correctly, but only 40% when the view is cluttered. Wrong answers vary with density accordingly, but the analysis is not significant (p=,055). The amount of undecided answers is correlated to density levels (p=,000). Users were undecided for only 1% of the trials with no density, as opposed to 40% with the cluttered one.

The experiment shows that H1 was confirmed: the level of density negatively impacts the effectiveness at tracking marks. In the next section, we propose a new rotation design that counters the effect of density.

3 Experiment 2: Improving 3D Rotation Perception and Perception of Relative Arrangement

If designers of visualization software want to make rotation effective at tracking marks in dense scenes, they have to provide specific features that counter the density effect. Our proposal is to set the rotation axis on the graphical mark of interest. We called this 3D-rotation a *focus-centred* rotation.

The experiments we show in this section are not about testing if focus-centred rotation is better than screen-centred rotation at tracking marks. However, perceiving arrangement of marks (i.e. understanding spatial relationship between marks) relies on correct tracking; considering experiment 1 results, we already know that density hinders perceiving arrangement. Thus, these experiments aim to show that a focus-centred 3D rotation helps better the perception of relative arrangement than a screen-centred 3D rotation. To summarize, the assumptions we make are: (H1) Focus-centred rotation provides a better accuracy than screen-centred rotation (H2) Density levels have a negative impact on accuracy; (H3) Density levels have less of a negative impact on accuracy with focus-centred rotation compared to screen-centred rotation.

The experimental apparatus and participants are the same as in the first experiment. In this experiment, two trajectories are displayed as lines. When seen in the top view, lines cross. However, one trajectory stays at the same altitude, whereas the other one changes its altitude. The user has to figure out if the stationary trajectory goes *under* or *above* the other one. This task requires the user to focus on the trajectories of interest, and to perceive the relative arrangement of the lines. Participants first see a top view with two highlighted crossing trajectories in red (the surrounding trajectories are green). After 4 seconds, the colour of the specific trajectories turns back to green, and the 3D rotation from the top to the vertical view is played. At the end of the transition, participants were asked to answer the question: "does the stationary trajectory go under the trajectory that changes altitude?" with "yes", "no", or "I don't know" We used three different density levels (empty, medium with 20 trajectories, high with 40 trajectories), 4 profiles of trajectories, 4 trajectory locations (lop-left, top-right, bottom-left, bottom-right), and 2 types of transition (screen-centred, focused-centred rotation). Since we did not repeat measures, users performed 96 trials each.

We first analyzed the global effect of rotation types without taking into account the context. The rotation types have a significant effect on correct answers: F(1, 10)=12,224, p=,00576 and on undecided answers: F(1, 10)=7,8955, p=,01848. Since the number of bad answers is very low, the analysis does not show a significant effect on the wrong answers: F(1, 10)=1,5432, p=,24248. This analysis shows that without taking into account the context density, the Focus-centered rotation improves the rotation efficiency to understand the data structure (H1).

We also assessed the impact of the context with repeated measures ANOVA using contexts (Empty, Medium, and Heavy) and the type of rotation (Screen-centered, Focus-centered) factors. For the participant's correct answers, we have found that there is a main effect of rotation type: F(2,20) = 12,224 p=,005762, and a main effect of context : F(2,20)=26,395 p=,000. We also found an interaction between density and rotation type (F(2,20)=6,754 p=,005740). These results show that density has less of an impact on correctness of answers with focus-centered rotation (from 99% to 90% vs from 99% to 66% for screen-centered rotation).

For wrong answers, we have found that there is no main effect. Results were not found significant for the rotation type effect (F(2,20)= 1,53918 p=,243058) and the density effect(F(2,20)= 1,51103 p=,244827). We did not found an interaction between density and rotation type (F(2,20)=0,61053 p=,0552848). These results are not significant because participants gave very few wrong answers. This suggests that both transitions are not misleading. For the participants' undecided answers, we have found that there is a main effect of rotation type (F(2,20)=7,89552 p=,018478) and a main effect of context (F(2,20) = 13,60981 p=,000186). We also found an interaction between density and rotation type (F(2,20) = 9,25000 p=,001431).

We can conclude that Focus-centered rotation minimizes the indecision of participants when density increases (for heavy density, 8% vs 28% undecided answers). Our analysis shows that the type of rotation has a significant effect on the user's ability to understand the data structure (H2) and that the impact of the context density is reduced with the Focus-centered rotation (H3). Hence the Focus-centered rotation is more efficient than the Screen-centered rotation.

4 Conclusion and Design Perspective

We proposed that rotation-based transition help in tracking marks, and perceiving their relative arrangement. We showed that users can track marks with a rotation, but that density hinders it. We also showed that rotation enables users to perceive relative arrangement, and that focus-centred rotation improves perception of relative arrangement compared to a screen-centred rotation. As stated before, this paper is about perception of rotation for visualization, and not about the interaction with the rotation. Now that we know that perception is better with a focus-centred rotation, we need to enable the user to rotate around this axis instead of the centre of the screen. Of course, in contrast to experiment 2, the system does not know which marks the user currently analyses. Hence, we are currently working on the design of an interaction to set the location of the axis. Solutions based on selecting an object are not suitable, since our scenes are very dense and cluttered, which makes pointing a difficult task.

References

1. Elmqvist, N., Dragicevic, P., Fekete, J.-D.: Rolling the dice: Multidimensional Visual Exploration using Scatterplot Matrix Navigation Visualization. IEEE Trans. On Visualization and Computer Graphics 14(6), 1148–1539 (2008)
2. Hurter, C., Conversy, C., Tissoires, B.: FromDaDy: Spreading Aircraft Trajectories Across Views to Support Iterative Queries. IEEE Transactions on Visualization and Computer Graphics 15(6), 1017–1024 (2009)
3. Robertson, G., Cameron, K., Czerwinski, M., Robbins, D.: Animated visualization of multiple intersecting hierarchies. Information Visualization 1(1), 50–65 (2002)

HTML 5 Support for an Accessible User-Video-Interaction on the Web

Lourdes Moreno, Paloma Martínez, Ana Iglesias, and María Gonzalez

LaBDA Group, Computer Science Department, Universidad Carlos III de Madrid
{lmoreno,pmf,aiglesia,mgonza1}@inf.uc3m.es

Abstract. Multimedia content covers the Web, and we should provide access to all people. For this reason, it is very important to take into account accessibility requirements in the player to avoid barriers and to ensure access to this multimedia content as well as their resources. One of the most frequent barriers is the technological obstacle: the necessity for the user to install the required plug-ins in to order to access video. The new standard HTML5 provides a solution to this problem. However, it does not fully support accessibility requirements of W3C standards, including WCAG and interaction requirement of UAAG. This paper[1] introduces an overall study of this new standard in relation to accessibility requirements for the players as well as an accessible HTML5 Media Player.

Keywords: Web accessibility, video HTML 5, user agent, media players.

1 Introduction

The growth and access of rich and interactive contents such as video and audio have long filled the Web. Multimedia content must be accessible for people with disabilities according to standards like the Web Content Accessibility Guidelines (WCAG) [1] of Web Accessibility Initiative (WAI) [2]. Additionally, multimedia accessibility content on the Web requires that a particular chain of essential, interdependent, and accessible components [3] should be taken into account for user agents. Specifically, media players should enable the delivery of accessible multimedia content and therefore a friendly user-video interaction would be a success. Furthermore, it is imperative that Web designers are familiar with the User Agent Accessibility Guidelines (UAAG) [4] of WAI as well as the existing media players capable of the accessible delivery of multimedia content.

The UAAG explains how to make user agents – including Web browsers, media players, and assistive technologies – accessible for people with disabilities, and particularly, how to increase the accessibility of Web content. Related to multimedia contents, there are accessibility requirements that must be taken into account respect to audio and video on the Web. In general, media players must ensure that all audio

[1] This research work is supported by the Research Network MAVIR (S2009/TIC-1542 (www.mavir.net/)), GEMMA (TSI-020302-2010-141) and SAGAS (TSI-020100-2010-184) research projects.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 535–539, 2011.

and video controls are accessible via keyboard and they should be accessed via screen readers. On the other hand, according the Guideline 1.2 of WCAG 2.0, media players must provide alternatives for time-based media and the media content must be accompanied by media alternatives as caption (or subtitles for deaf people), audio description, sign language, etc. in synchronized media. Moreover, according UAAG 2.0 guidelines, players should provide support for these media alternatives.

Previous studies [5] demonstrated that current media players are not sufficiently accessible. Most of the online videos are delivered via Adobe Flash[2]-based in-page video players. Flash has an excellent compression system that can deliver high-fidelity audio and high-resolution video without taxing bandwidth. However, devices highly used for delivering video such as Apple[3] - iPad and Apple - iPhone do not have support for this format.

As Universal alternative for solving accessibility problems appeared HTML5 [6]. This marked language allows video delivery on the Web to be independent from used user devices, introducing the **<audio>** and **<video>** elements, which enables web browsers to natively support media.

2 Study of HTML 5 Support for Accessibility Standards

The W3C is currently working on HTML 5, which will replace XHTML standards. HTML5 adds implicit semantic information, defines the basis of the document-oriented model (DOM), and presents a better structure. Regional landmarks on a webpage will have their own tags so that user agents can recognize them, which in turn will allow Web search robots gather more accurate information, letting authors embed accessibility multimedia elements natively (using **<audio>** and **<video>** tags), obviating the need for plug-ins.

HTML5 offers access in the embedded media player to be a huge step forward. The new standard introduces the latest commands (such as **<video>** and **<audio>),** which can create and label controls letting keyboard shortcuts to access them, and screen readers to tell the user which controls are available. Some HTML5 elements provide support for some UAAG 2.0 guidelines. However, the inclusion of caption and audio description, have not yet been included in the new standard because this is still under development. Problems have been detected in the current versions of most browsers in the support of **<video>** and **<audio>** labels[4]. Unfortunately, none of the current browser implementations are fully accessible by keyboard and screen-reader[5]. Others complications arise because the various browsers can't agree on what video codecs they're going to support, so actually using HTML5 **<video>** on the web is currently much more difficult than it should be[6].

---

[2] 2011 Adobe Systems Incorporated, `http://www.adobe.com/`

[3] Apple, `http://www.apple.com/es/`

[4] New HTML5 feature accessibility support in Windows Browsers, `http://html5accessibility.com/`

[5] Keyboard and Screen-reader Accessibility `http://terrillthompson.blogspot.com/2010_08_01_archive.html`

[6] Dive Into HTML5, What Works on the Web `http://diveintohtml5.org/video.html#what-works`

The current version of HTML5 provides some playback controls for the video: play/pause, full screen toggle, volume and audio element controller toolbar. These controls are very basic, not allowing the user to have full control of the video. Following WCAG 2.0 and UAAG 2.0 guidelines, these native controls of HTML 5 are not enough. It must have the following controls: end (stop), caption on/off, search captions for text strings and select caption language if closed captioning is available, audio description on/off, rewind/forward seconds, volume up/ down, screen reader full access and keyboard full access of controls.

In this work, we have developed an accessible HTML5 Media Player[7] as it is shown in the next section.

3 Accessible HTML5 Media Player

We have developed an accessible HTML5 Media Player, in order to build the mentioned controls as complement HTML5 support for interaction requirements included in UAAG 2.0, the use of JavaScript or Flash technology is required. We have chosen JavaScript because it does not require an external plug-in. Furthermore we have carried out different codifications of the same video to assure that the video can be played on all browsers. Currently there are three media formats that are under discussion of the different user agents: MP4, OGV, and WebM.

The attributes are included within **<video>** tag (see Figure 1), some of their values are called for JavaScript functions in order to achieve a high level of accessibility in the interaction with the video based on the standard UAAG 2.0. Some of the attributes are "**controls**" to load the native controls on devices with JavaScript disabled, and "**onKeyDown**" to control the video from the keyboard, so the user can use it for pause the video, stop it, show captions, etc. This function has been implemented with JavaScript.

```
<video width="660" id="player" onKeyDown="KeyControls(event)" onclick="repro()">
          <source src="21.mp4" type="video/mp4" />
          <source src="21.theora.ogv" type="video/ogg" />
          <source src="21.webm" type="video/webm">
</video>
```

Fig. 1. HTML5 code of **<video>**

Controls have been incorporated through JavaScript, included in a separate <div> for the rest of the code, as it is shown in Figure 2.

```
<div id="controls" onMouseOut="HideControls()"onMouseOver="ShowControls()"
onKeyDown="KeyControls(event)">
```

Fig. 2. HTML5 code for controls

Because its complexity, it is not possible to describe in detail the code of each of the additional controls. Each of them has been implemented and its result is shown in Figure 3.

---

[7] Available in `http://labda.inf.uc3m.es/LourdesPlayer/`

In order to add a caption in the player in HTML 5, an XML file was created with it. From this XML, file subtitles are being extracted and incorporated in a synchronized way through JavaScript and DOM is being modified.

In the case of access from devices with disabled JavaScript or CSS, the player will only appear with the HTML 5 native controls, allowing users access the video content. Moreover, to increase accessibility Keyboard Shortcuts have been defined to access the same functionality implemented using JavaScript technology and video transcriptions have also been provided. Finally, following UAAG requirements, which make reference to the usability, an alternative text have been associated for each of the player controls and the player offers a help guide, describing the keys associated with the keyboard shortcuts among others.

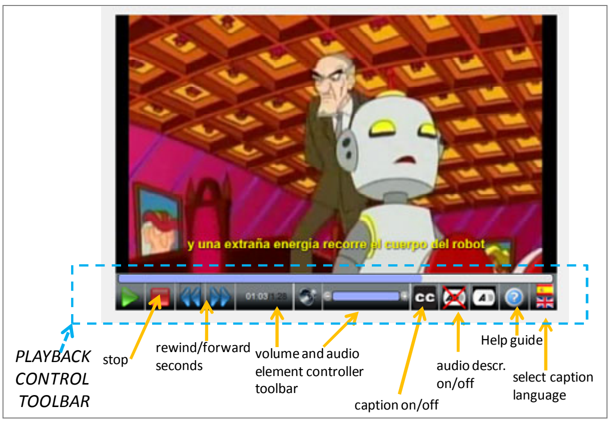


Fig. 3. Accessible HTML5 Media Player

4 Conclusions

HTML 5 represents a step forward in the multimedia history on the Web, but there is still a lot of work to do in order to include accessibility requirements on user interaction, such as those that were done explicitly in this work based on W3C standard such as WCAG and UAAG. This issue is aggravated by the fragmentation between platforms. The question is posed, how should we deliver video, in Flash or HTML5? Currently, Flash supports more advanced interactions, but it presents some accessibility barriers. On the other hand, HTML5 provides accessible solutions, but it is still been developed and not all browsers implement it yet. Therefore, we suggest right now the smart strategy of implementing websites using both approaches, i.e.: approaches with accessible Flash players and with combination of HTML5 and JavaScript technology solution as it has been shown in this work.

References

1. W3C, WAI, WCAG (2011), http://www.w3.org/WAI/intro/wcag.php
2. W3C, Web Accessibility Initiative (WAI) (2011), http://www.w3.org/WAI/
3. Moreno, L., Martínez, P., Ruiz-Mezcua, B.: Disability Standards for Multimedia on the Web. IEEE Multimedia 15(4) (2008)
4. W3C, WAI, UAAG (2011), http://www.w3.org/WAI/intro/uaag.php
5. Moreno, L., Gonzalez, M., Martínez, P., Iglesias, A.: A Study of Accessibility Requirements for Media Players on the Web. In: Stephanidis, C. (ed.) HCII 2011 and UAHCI 2011, Part I. LNCS, vol. 6765, pp. 249–257. Springer, Heidelberg (2011)
6. W3C. HTML5. A vocabulary and associated APIs for HTML and XHTML. Working Draft (April 05, 2011), http://dev.w3.org/html5/spec/Overview.html

Improving the Flexibility of Model Transformations in the Model-Based Development of Interactive Systems

Christian Wiehr[1], Nathalie Aquino[2], Kai Breiner[1], Marc Seissler[3], and Gerrit Meixner[4]

[1] University of Kaiserslautern, Software Engineering Research Group, Gottlieb-Daimler Str. 42, 67663 Kaiserslautern, Germany
{c_wiehr,Breiner}@cs.uni-kl.de
[2] Research Center on Software Production Methods, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain
naquino@pros.upv.es
[3] University of Kaiserslautern, Center for Human-Machine-Interaction, Gottlieb-Daimler Str. 42, 67663 Kaiserslautern, Germany
Marc.Seissler@mv.uni-kl.de
[4] German Research Center for Artificial Intelligence (DFKI), Trippstadter Str. 122, 67663 Kaiserslautern, Germany
Gerrit.Meixner@dfki.de

Abstract. This paper presents an approach that adds flexibility in the varieties of user interfaces that can be generated by processes of model-based user interface development. This approach is used at design time. Ideas from this approach have been extended for use at runtime and have been applied to SmartMote, a universal interaction device for industrial environments.

1 Introduction

Developers are faced with the problem of having to build user interfaces (UIs) for a plethora of target devices and usage situations. Model-based user interface development (MBUID) methodologies promise to reduce the complexity of this problem by leveraging different layers of abstraction, the respective models for expressing aspects of UIs on these levels, and transformation tools for the development and semi-automatic generation of UIs. Frameworks, such as the Cameleon Reference Framework (CRF) [6], have been proposed, which define the different abstraction layers and models to be used for the systematic, user-centered design of multi-target and context-aware UIs.

However, the model-based development of context-aware and runtime adaptive UIs still presents relevant challenges regarding model transformations. When an approach uses a transformation language/model to specify transformations, we say that it has *explicit* transformation logic. In concepts like DynaMo-AID [7] and MASP [4], model transformations and adaptations during runtime are implemented in the underlying renderers, that is, using *implicit* transformation logic. While the MARIA language [9] uses XSLT for the model transformations, it is unclear how the model adaptation during runtime is specified in this concept.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 540–543, 2011.

Approaches with implicit transformation logic have limitations regarding the diversity of user interfaces that can be generated and they might require manual modifications on the generated code in order to deal with UI requirements that are not supported by the automated transformation process. Approaches with explicit transformation logic can be considered complex to be used by UI designers and there is a lack of suitable support for manipulation of models at runtime. Therefore, new transformation mechanisms are needed that support the explicit specification of model transformations and manipulations in order to increase the flexibility of today's engineering processes for context-aware UIs.

In Section 2, we present the Transformation Templates approach, which serves as the initial mapping concept for our transformation approach. In Section 3, we extend our runtime generation approach to improve the flexibility of its transformation process. In Section 4, we discuss results and conclude.

2 Flexibility at Design Time: Transformation Templates

A Transformation Template (TT) [2] aims to explicitly specify the structure, layout, and style of a UI according to the preferences and requirements of the end users as well as in line with the different hardware and software computing platforms and environments in which the UI will be used, i.e., different contexts of use.

A *parameter type* represents a design or presentation option. Defining a parameter type subsumes specifying the types of elements of a UI model that are affected by it, as well as its *value type*, which refers to a specific data type or to an enumeration of possible values. A parameter type can be implemented for different contexts of use and usability guidelines are provided in order to support the selection of suitable values by UI designers in different situations. A TT gathers a set of parameters for a specific context of use. Each *parameter* corresponds to a parameter type and has both a value and a selector. The *value* of a parameter corresponds to a possible value of the corresponding parameter type. A *selector* delimits the set of UI elements that are affected by the value of a parameter. We have defined different types of selectors that allow the designer to choose different sets of UI elements.

TTs are used to parameterize model-to-model or model-to-code transformations. A model compiler takes the source UI model and a TT as input. The values and selectors of the parameters of the TT specify how to transform the source UI model into the target UI model. It is important to note that TTs do not replace any implicit transformation logic or explicit transformation languages; instead, they provide a higher-level tier for UI designers to easily specify UI transformations at design time.

TTs add flexibility in MBUID approaches because they externalize the design knowledge and presentation guidelines and make them customizable according to the characteristics of the project being carried out. TTs can then be reused in other projects with similar characteristics. Furthermore, TTs aim to diversify the kinds of UIs that a MBUID approach can generate.

3 Using TTs to Increase Flexibility at Runtime

In our MBUID approach for runtime adaptive systems, three core models can be identified (see Fig. 1): 1) the Useware Markup Language (useML) [8], which is used to structure the user's tasks; 2) the Dialog and Interface Specification Language (DISL) [3], which is used for describing the dialog behavior of the UI; and 3) the User Interface Markup Language (UIML) [1], which is used to define how the content is presented to the user in terms of concrete interaction objects and their layout.

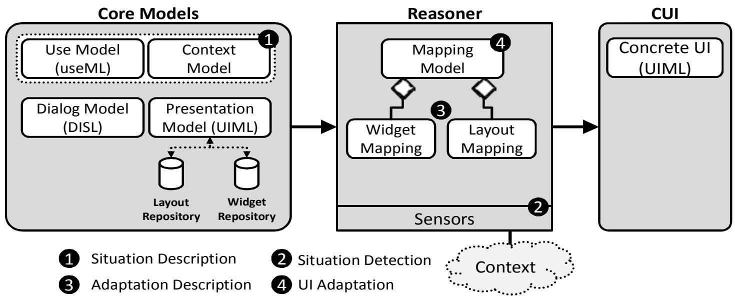


Fig. 1. Transformation process in our MBUID approach for runtime adaptive systems

These models together with runtime transformation specifications coded in the generator software (implicit transformations) were enough to obtain functional but basic UIs for runtime adaptive systems. In order to improve the flexibility of model transformations in our approach, we took the idea of the TTs as an initial mapping concept and we extended it to refer to dynamic, runtime model data, so the UI can be automatically generated and react to adjusted models.

Therefore, a *context model* has been integrated to provide access to static context information (such as information about the user or environment that may have a direct influence on the interaction) (see Fig. 1). A *mapping model* with mapping rules was also integrated as an extension to TTs (see Fig. 1) to refer to model data in addition to the fixed values of the TTs. It is composed of *layout mappings*, which allow the structure of the UI to be defined according to a hierarchical structure of UI containers and widgets; and *widget mappings*, which allow the mappings from abstract interaction objects to concrete widgets to be described. Further, in order to support the transformation process in terms of reuse and performance, a repository containing frequently occurring widgets (*widget repository*) and container-widget configurations (*layout repository*) was integrated (see Fig. 1), that can be extended easily. Both, the containers and the widgets are specified using UIML – featuring parameters to customize their presentation. At runtime, these parameters are set according to the mapping rules, using either fixed values or references to model data. All of the transformation specification previously coded in the generator software can be formalized using the mappings, which guarantee by definition that the generation will be successful. In order to demonstrate the feasibility of our approach, we developed a functional prototype as an extension to the SmartMote approach [5].

4 Discussion and Conclusion

In this paper, we presented a new mapping concept for improving the flexibility of an approach for developing context-aware UIs on the basis of UI models according to the levels of abstraction in the CRF.

For the development of such UIs, the underlying models have to be automatically manipulated during runtime. The TT approach can serve as the underlying concept, but it had to be extended. An extended TT concept was developed that supports the interlinking and mapping of different models during runtime. Mapping rules can use model values as input and manipulate the UI generation process during runtime. To test the feasibility of this concept, a first prototype was developed.

Acknowledgments. This work has been developed with the support of MICINN under the project PROS-Req (TIN2010-19130-C02-02), and GVA under the project ORCA (PROMETEO/2009/015) and the BFPI/2008/209 grant, and co-financed with ERDF. We also acknowledge the support of the ITEA2 Call 3 UsiXML project under reference 20080026. Parts of the presented work are result of the GaBi project funded by the German Research Foundation (DFG) which is part of the AmSys research focus at the University of Kaiserslautern funded by the Research Initiate Rhineland-Palatinate.

References

1. Abrams, M., Phanouriou, C., Batongbacal, A.L., Williams, S.M., Shuster, J.E.: UIML: An Appliance-Independent XML User Interface Language. 31, 1695–1708 (1999)
2. Aquino, N., Vanderdonckt, J., Pastor, O.: Transformation templates: adding flexibility to model-driven engineering of user interfaces. In: Biryukov, A., Gong, G., Stinson, D.R. (eds.) SAC 2010. LNCS, vol. 6544, pp. 1195–1202. Springer, Heidelberg (2011)
3. Bleul, S., Schäfer, R., Müller, W.: Multimodal Dialog Description for Mobile Devices. In: Gehalten auf der Workshop on XML-based User Interface Description Languages at AVI 2004, Gallipoli, Italy (2004)
4. Blumendorf, M., Feuerstack, S., Albayrak, S.: Multimodal user interfaces for smart environments: the multi-access service platform. In: Proceedings of the Working Conference on Advanced Visual Interfaces, pp. 478–479. ACM, Napoli (2008)
5. Breiner, K., et al.: Automatic adaptation of user workflows within model-based user interface generation during runtime on the example of the SmartMote. In: Proceedings of the 15th International Conference on Human-Computer Interaction (HCII 2010), Orlando, FL (2011)
6. Calvary, G., et al.: A unifying reference framework for multi-target user interfaces. Interact Comput. 15(3), 289–308 (2003)
7. Clerckx, T., Luyten, K., Coninx, K.: DynaMo-AID: a Design Process and a Runtime Architecture for Dynamic Model-Based User Interface Development. In: The 9th IFIP Working Conference on Engineering for Human-Computer Interaction Jointly with the 11th International Workshop on Design, Specification and Verification of Interactive Systems, pp. 11–13. Springer, Heidelberg (2004)
8. Meixner, G., Seissler, M., Breiner, K.: Model-Driven Useware Engineering. In: Hußmann, H., Meixner, G., Zühlke, D. (eds.) Model-Driven Development of Advanced User Interfaces, pp. 1–26. Springer, Heidelberg (2011)
9. Paternó, F., Santoro, C., Spano, L.D.: MARIA: A universal, declarative, multiple abstraction-level language for service-oriented applications in ubiquitous environments. ACM Trans. Comput.-Hum. Interact. 16, 1–30 (2009)

In Two Minds about Usability? Rationality and Intuition in Usability Evaluations

Volker Thoma and Elliott P. White

University of East London, Romford Road, London E15 4LZ

Abstract. Usability ratings of a university website by 60 students were analysed together with participant's self-ratings of their cognitive style. The degree of users' "rational" as well as their "intuitive" style correlated with usability evaluation scores. In particular, self-reported rational ability was connected with evaluations of Controllability, intuitive ability was related to Helpfulness scores of the interface. Thinking style significantly affects usability ratings (explaining over 9% of the ratings' variation), which has implications for evaluations across user groups.

1 Introduction

User-centred design relies on a variety of methods and techniques which aim to involve potential users at different stages of product development. A common component of such projects often is a summative evaluation that aims to measure "ease of use" [1, 2]. The results of user evaluations affect decision making in IT projects, but the factors influencing the users' judgments are rarely analysed, even if they are subdivided into "user groups" according to socio-economic criteria or persona characteristics [3]. Furthermore, only recently researchers and practitioners have embraced the concept of "affect" or "attractiveness" to account for "hedonic" or "fun" aspects of use [4]. However, modern research on human judgment, decision making, and social psychology has identified a great number of factors that may systematically influence preferences or ratings. For example, Slovic [5] identified the *affect heuristic* - a positive or negative feeling experienced by the evaluator–informs their preferences between items or images.

There is growing evidence that human reasoning and decision making is the product of two distinct cognitive systems. Researchers share the labels 'System 1' and 'System 2' for these two systems [6], although they are more commonly referred to as "intuition" and "rationality". System 1 is a fast, effort-free collection of autonomous processes that are susceptible to emotional stimulation. System 2 is characterised by slow, deliberative, rule-based reasoning [6]. Humans seem to vary in their propensity for intuitive versus reflective thinking in their daily judgments. Thus, cognitive style, also called thinking style, refers to a person's preferred way of processing information and the way he or she perceives, conceptualises, and judges information [6]. One of the earliest and most reliable methods to investigate cognitive style is the Rationality and Experientiality Inventory (REI) developed by Epstein et al [7]. The subscales of this questionnaire correspond to self-rated ability and preference of utilizing System 1 and System 2 respectively. Rationality and Experientiality have been found to be

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 544–547, 2011.

orthogonal constructs [7] and are dissimilar to cognitive abilities because they do not describe peak performance, but they do influence attitudes and social interaction.

The aim of the current study is to test whether individual differences in "Experientiality" and "Rationality" may influence user evaluations, and whether aspects of usability ratings depend more on intuition rather than reflective thought. We used the WAMMI (**W**ebsite **A**nalysis and **M**easure**M**ent **I**nventory) [2] (after considering SUMI, QUIS, SUS [1]) to elicit usability ratings of a website, because of its widespread use, and because it employs a separate subscale of Attractiveness (A). Our predictions were that a) WAMMI scores are influenced by individual differences in thinking as measured by the REI and b) that some WAMMI subscales would correlate more strongly with rationality than others (A), as people would rely more on system 1 processes (i.e., automatic, implicit visual evaluations) in this case.

2. A Web Site Evaluation and the Role of Intuition and Reflexivity

Participants, Measures and Procedure

Sixty students (40 females, age range between 18 and 43, mean = 24.45) in a modern UK university were approached in one of the campus computer centres and invited to take part in a study to evaluate the university's website.

Once participants agreed they were asked to visit to the university's homepage. They were not required to perform any tasks but simply asked whether they had used the website before. If so, participants completed the two questionnaires after brief inspection of the website. The WAMMI was used to elicit ratings regarding the usability of the website. The WAMMI is a 20 item questionnaire in which participants rate their strength of approval for statements like "The pages on this web site are very attractive" and "This web site has some annoying features", on a 1 – 5 scale. Data from this scale were scored by the WAMMI's authors using a proprietary algorithm. Results are provided on five sub-scales: Learnability (L), Helpfulness (H), Efficiency (EC), Controllability (C), and Attractiveness (A), and a combined global usability score (G). After that participants filled in the Rational-Experiential Inventory-short form, a 24 item version of the original scale [7]. Participants rate the strength of their agreement with statements like "I have a logical mind" (Rational ability) and "I believe in trusting my hunches" (Experiential favourability), on a 1-5 scale. These items combine into four subscales: Rational Ability (RA), Rational Favorability (RF), Experiential Ability (EA), and Experiential Favorability (EF), which can be summed up for overall 'Rationality' (R) and 'Experientiality' (E) scores.

Results

Pearson correlation coefficients were calculated. Table 1 shows the pattern of correlations and level of significance. Global WAMMI scores correlated significantly with RA and Rationality, as well as with Experientiality. The WAMMI subscale for Controllability and Efficiency correlated with RA and Rationality, whereas Helpfulness correlated with RA, EA and Experientiality. Learnability correlated with EF only. Interestingly, against our prediction the subscale of Attractiveness did not correlate with experiential scales, but it did correlate with RA.

Table 1. Correlations of WAMMI and REI scales

			WAMMI					REI					
		A	**C**	**E**	**H**	**L**	**G**	**RA**	**RF**	**R**	**EA**	**EF**	**E**
WAMMI	**Attractiveness**		**.59****	**.60****	**.60****	**.54****	**.83****	**.28***	.11	.23	.186	.05	.15
	Controllability			**.67****	**.56****	**.50****	**.84****	**.39****	.24	**.37***	.17	.16	.20
	Efficiency				**.48****	**.39****	**.79****	**.31***	.23	**.33***	.06	.10	.09
	Helpfulness					**.52****	**.8****	**.31***	.10	.25	**.41****	.21	**.39****
	Learnability						**.72****	.08	.15	.14	.09	**.34****	.25
	Global							**.35****	.21	**.33****	.23	.21	**.27***

**= significant at the 0.01 level, *=significant at the 0.05 level.

In order to test whether individual differences in cognitive style can explain variations in how people rate the usability of this interface, we ran simple linear regression analyses with the REI subscales as predictors and the WAMMI scores as criterion. In the first analysis with the global WAMMI score as predictor, the ANOVA was marginally significant, $F(4, 55) = 2.51$, $p = .052$, meaning that the REI model predicted the global WAMMI scores to a significant degree ($R=.39$). However, only RA approached significance as a predictor (Beta = .26, $t = 1.850$, $p = .07$) all other predictors were $t =< 1.07$. Thus, the REI model explained 9% of the variance in WAMMI-G scores. For the analyses with a WAMMI subscore (A, EC, L) as criterion we obtained no significant model, all $p > .1$. However, for Controllability the REI model significantly predicted the WAMMI scores, $F(4, 55) = 2.70$, $p < .05$ (adjusted $R^2= 0.10$), again with RA as the only significant predictor ($t = 2.29$, $p < .05$). The model reliably predicted Helpfulness, $F(4, 55) = 3.69$, $p< .05$, (adjusted $R^2= 0.15$), this time only with EA explaining a significant part of the variance ($t = 2.62$, $p < .05$).

3 Conclusions

We asked whether usability measures could be differentially influenced by individual differences in cognitive style. There were two important findings. First, usability scores were higher the more people believed in their thinking ability: both "rational" and "experiential" scores correlated with WAMMI ratings. However, only rational ability self-ratings significantly predicted variations in WAMMI scores in a multiple regression model. It is outside the scope of the study to determine whether this means that people who feel they can rely on their (rational) thinking style in general feel more positive ("in control") about an interface, or whether this reflects overconfidence in participants about their abilities which may result in a sense of mastery of the interface. Whatever the interpretation of this effect, if confirmed this result would mean that comparing summative usability evaluations across different populations may be affected by differences in thinking style. For example, some Asian populations are found to be less overconfident in their judgments than Westerners and prefer a more global style of information processing [8, 9]. Hence, cross-cultural user research may be susceptible to thinking style.

The second result is more complex: self-belief in rational ability positively affects the experience with an interface, in particular in terms of its controllability. One explanation is that people who score high on RA are able to cope with poor interactive design features and feel less affected by it. For example, they may work out how to navigate even a poorly organised website and attribute this relative ease as inherent to the website - thereby inflating the Controllability scores. On the other hand, Helpfulness (but not Attractiveness) scores correlated highly with a global “Experiential” style. The reasons for this are probably complex, but it is possible that the more “intuitive” participants relied successfully on processes which are unconscious and work automatically. Compared to the task of deciding whether a website is easy to control by recalling specific episodes of usage, judgments on Helpfulness may be harder to break down into distinct instances. As the experiential system operates in an automatic and holistic manner, users that score high on Experiential Ability may have simply a better mental “toolset” for making such a more global judgment, resulting in higher scores. In contrast, users may rely on rational criteria when asked to judge Attractiveness [10].

In conclusion, the current study provides evidence for a role of individual thinking styles in usability evaluations. Further research would need to establish - for example by using “verbal protocols” and different memory tests - whether user ratings can indeed be traced to reflective or intuitive processes .

References

[1] Holzinger, A.: Usability Engineering for Software Developers. Commun. ACM 48(1), 71–74 (2005)

[2] Kirakowski, J., Claridge, N., Whitehand, R.: Human Centered Measures of Success in Web Site Design. In: Proceedings of the Fourth Conference on Human Factors & the Web, Basking Ridge, NJ (1998), http://www.ucc.ie/hfrg/questionnaires/wammi/research.html

[3] Thoma, V., Williams, B.: Developing and Validating Personas in e-Commerce: A Heuristic Approach. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5727, pp. 524–527. Springer, Heidelberg (2009)

[4] Hassenzahl, M.: The interplay of beauty, goodness, and usability in interactive products. Int. J. Hum-Comput. Int. 19, 319–349 (2004)

[5] Slovic, P., Finucane, M.L., Peters, E., MacGregor, D.G.: The affect heuristic. In: Gilovich, T., Griffin, D.W., Kahneman, D. (eds.) Heuristics and biases: The psychology of Intuitive Judgment, pp. 397–420. Cambridge University Press, Cambridge (2002)

[6] Evans, J.: In two minds: dual-process accounts of reasoning. Trends Cogn. Sci. 7, 454–459 (2003)

[7] Epstein, S., Pacini, R., Denes-Raj, V., Heier, H.: Individual differences in intuitive-experiential and analytical-rational thinking styles. J. Pers. Soc. Psychol. 71, 390–405 (1996)

[8] Lee, J.W., Yates, J.F., Shinotsuka, H., Singh, R., Onglatco, M.L.U., Yen, N.S., Gupta, M., Bhatnagar, D.: Cross-national differences in overconfidence. Asian J. Psy. 1, 63–69 (1995)

[9] Zhu, Y., Ronis, D.L., Wang, D.F., Shinotsuka, H., Toda, W.: Probability judgment accuracy: China, Japan, and the United States. Organ. Behav. Hum. 43, 145–171 (1989)

[10] Leder, H., Belke, B., Oeberst, A., Augustin, D.: A model of aesthetic appreciation and aesthetic judgments. Brit. J. Psychol. 95, 489–508 (2004)

Influence of Web Content Management Systems in Web Content Accessibility

Juan Miguel López[1], Afra Pascual[2], Llucia Masip[2], Toni Granollers[2], and Xavier Cardet[2]

[1] Department of Computer Languages and Systems, School of Engineering, University of the Basque Country, Nieves Cano 12, 01006 Vitoria-Gasteiz, Spain
juanmiguel.lopez@ehu.es
[2] GRIHO research group, Polytechnic School, University of Lleida, Jaume II 69, 25001 Lleida, Spain
{apascual,lluciamaar,tonig}@diei.udl.cat, xcardet@alumnes.udl.cat

Abstract. Web Content Management Systems (CMS) are traditionally used in institutions to allow web content management by people without technical skills. This study intends to check the influence of the CMS in the accessibility of the contents they handle. First, an accessibility analysis of six widely used CMS is performed base don the Authoring Tool Accessibility Guidelines (ATAG). Second, an accessibility analysis of a series of city council web pages managed by abovementioned CMS by using Web Content Accessibility Guidelines (WCAG). Results of the study show that although web pages managed by the CMS with a better degree of ATAG fulfilment provide better accessibility, there is no direct correlation between results obtained in both evaluations. Information about what aspects cause accessibility errors in the CMS and the impact of such aspects in accessible web content management is also provided.

Keywords: Web Content Management Systems, Authoring Tool Accessibility Guidelines, Web Content Accessibility Guidelines.

1 Introduction

Currently, there are powerful legal requirements to take accessibility into account as a requirement for developing web sites. For instance, these kinds of regulations have been performed on European Union countries [1]. It is important to note that these legal frameworks do not only affect public administration, but also corporations that develop software or provide services for their use in civil service.

Nowadays there is a wide range of institutions that use CMS to manage their web content. They are especially useful as they allow users without technical skills to introduce eliminate or modify web content, but it does have some risks regarding accessibility [2] [3]. Furthermore, CMS generally do not generate accessible web content by default and allow users to perform few modifications to improve the

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 548–551, 2011.

accessibility of the web content they manage. These kinds of modifications are frequently insufficient to allow websites to be accessible for people with disabilities [4].

2 Study

Six CMS were selected for the study. The selection was based on the CMS to be open source, commonly used in civil service and general purpose systems. Selected CMS were: Plone, Joomla!, Typo3, EzPublish, OpenCMS and Drupal. First, an accessibility analysis of the CMS on their default installations was performed based on the Authoring Tool Accessibility Guidelines (ATAG) [5]. In order to complete the study, a web content accessibility analysis was performed on a sample of 90 worldwide town council web pages, 15 of them managed by each CMS. Web Content Accessibility Guidelines (WCAG) [6] were used for analysing the accessibility of each web page using the evaluation methodology established by the World Wide Web Consortium [7]. A total of three evaluators performed the study, each of them evaluating two CMS. Although it would have been desirable that every CMS would have been evaluated by more that one evaluator, the complexity of CMS led to estimate positively the fact that evaluators had a high level of knowledge regarding CMS to be evaluated. The results obtained by these evaluators were estimated to be more useful compared to the ones that people without such knowledge should provide. Evaluators also had deep knowledge on both accessibility guidelines used in this study.

Table 1. Quantitative results of the ATAG evaluation for every selected CMS

Priority	Drupal			EzPublish			Joomla!			OpenCMS			Plone			Typo3		
	C	Ac	I	C	Ac	I	C	Ac	I	C	Ac	I	C	Ac	I	C	Ac	I
1	3	2	3	3	0	5	2	1	5	1	1	6	3	3	2	2	1	5
2	1	0	6	2	0	5	0	4	3	1	1	5	1	5	1	1	1	5
3	1	0	4	1	0	4	0	2	3	0	0	5	1	3	1	0	1	4
Relative	0	0	7	1	0	6	0	2	5	0	0	7	3	1	3	0	2	5

Table 2. Sum of detected accessibility errors for all web pages of each analysed CMS

Priority	Drupal	EzPublish	Joomla!	OpenCMS	Plone	Typo3
1	20	26	32	83	248	39
2	341	167	566	1265	422	571
3	83	98	219	168	169	105

Table 3. Mean and standard deviation of detected accessibility errors for all web pages of each analysed CMS

Priority	Drupal		EzPublish		Joomla!		OpenCMS		Plone		Typo3	
	M	St	M	St	M	St	M	St	M	St	M	St
1	1,3	2,4	1,7	2,7	2,1	2,6	5,5	14,5	16,5	15,5	2,6	4,5
2	22,7	32,3	11,1	14,1	37,7	44,1	84,3	98,2	27,8	30,5	38,1	70,7
3	5,5	4,1	6,5	6,9	14,6	16,8	11,2	13,4	12	12	7,0	9,4

Table 1 summarizes the results obtained in the ATAG evaluation, grouped by CMS. Rows are organized by the type of ATAG priority. Columns are labelled as: “C” (indicates the number of guidelines that were correctly fulfilled), “Ac” (indicates the number of guidelines with almost all requirements satisfied or bugs need to be fixed) or I (indicates the number of guidelines that were not correctly fulfilled). Guidelines with relative priority were evaluates separately, although it must be taken into account that they can be interpreted as different priority depending on the case. Table 2 resumes the results obtained in performed web accessibility evaluations. Due to space constraints, results are expressed by means of the number of accessibility errors for the different analysed guidelines, grouped by priorities. It must be taken into account that the table sums up all errors located for the 15 web pages analysed per CMS. On the other hand, Table 3 shows the mean (“M”) and the standard deviation (“St”) of the accessibility errors detected for the 15 web pages analysed per CMS, ordered by priority.

3 Discussion and Conclusions

Regarding quantitative results corresponding to the six analysed CMS, Drupal and Plone show more complete features regarding accessibility in their default installation. Anyhow, none of them achieved an “A” level of compliance regarding ATAG fulfilment. One of the main issues arising from the study is that the web editors used by default in analysed CMS do not allow generating accessible content. In this sense, they must be reconfigured, be discarded or allow installing web content editors external to the CMS so users can generate accessible web content. Another relevant aspect is the need to change default CMS templates, as they cause web elements’ layout not to be accessible. The need to improve documentation regarding the accessibility that CMS provide is another aspect to assess.

Regarding the quantitative analysis of the web content accessibility of the sample of analysed web pages, it is remarkable the fact that a minimum number of web pages reaches even the “A” compliance level established by WCAG 1.0 [6], despite the legality regarding accessibility that analysed web pages should fulfil. Interestingly, it is worth noting that only 1 web page managed by Plone and 9 managed by Drupal achieve such compliance level. Accessibility errors found in these web pages are due to a series of reasons:

1. Most priority 1 errors were due to the lack of a proper alternative text in images (WCAG 1.1 checkpoint)
2. A large number of web pages used absolute measures instead of relative ones to position elements in web pages (WCAG 3.4 checkpoint).
3. Use of device-dependent events (associated only to mouse events by) and lack of redundancy for these cases.
4. The fact that default installation of most CMS (except for Plone and Drupal) did not provide by default a clear content and layout separation has led to accessibility errors in almost all the web paged managed by them (WCAG 3.3 and 11.2 checkpoints).

The results of CMS accessibility evaluation show no direct correlation with the results of web accessibility evaluation. For instance, Plone shows relatively good results regarding ATAG fulfilment, but poor results regarding the accessibility of web pages managed using this CMS compared to other CMS. However, web pages managed using Drupal and EzPublish show less accessibility errors compared to the rest.

This work intends to highlight the necessity of the CMS to manage adequately the accessibility of the contents they manage. One interesting conclusion derived from this work is that all analysed CMS can be configured so they can be accessible. In order to do so, it is necessary to modify internal CMS classes and templates that adjust web content in an accessible way. Besides, the use of web content editors external to the CMS can allow managing accessible web code. There is also the option of including plugins in all the CMS that can be used to manage concrete aspects and can be used as tools to improve accessibility. Anyhow, all these solutions require deep and exhaustive knowledge regarding each CMS. There are some resources that allow improving CMS accessibility [8] [9], but they do not provide a clear methodological approach in providing different concrete solutions. Furthermore, it is important to provide accessibility monitoring systems to check the accessibility of managed web content [10].

Acknowledgements. This work has been partially funded by Spanish Ministry of Science and Innovation through TIN2008-06228 and TIN2008-06596-C02-01 research projects. Llucia Masip is the recipient of a pre-doctoral fellowship from the University of Lleida.

References

1. Article 26 of the European Union Charter of Fundamental Rights (extracted on June 2011 from), `http://www.europarl.europa.eu/charter/pdf/text_en.pdf`
2. Burzagli, L., Billi, M., Gabbanini, F., Graziani, P., Palchetti, E.: The use of current content management systems for accessibility. In: Miesenberger, K., Klaus, J., Zagler, W.L., Burger, D. (eds.) ICCHP 2004. LNCS, vol. 3118, pp. 331–338. Springer, Heidelberg (2004)
3. Rainville-pitt, S., D'Amour, J.M.: Using a CMS to create fully accessible websites. In: W4A2007- ACM International Conference Proceeding Series (2007)
4. Schulz, M., Pieper, M.: Web Compliance Management: Barrier-Free Websites Just by Simply Pressing the Button? Accessibility and the Use of Content-Management-Systems. In: Stephanidis, C., Pieper, M. (eds.) ERCIM Ws UI4ALL 2006. LNCS, vol. 4397, pp. 419–426. Springer, Heidelberg (2007)
5. Treviranus, J., McCathieNevile, C., Jacobs, I., Richards, J.: Authoring Tool Accessibility Guidelines 1.0. (extracted on June 2011 from), `http://www.w3.org/TR/WAI-AUTOOLS/`
6. Chisolm, W., Vanderheiden, G., Jacobs, I.: Web Content Accessibility Guidelines 1.0. (extracted on June 2011 from), `http://www.w3.org/TR/WAI-WEBCONTENT/`
7. Evaluating Web Sites for Accessibility: Overview, (extracted on June 2011 from), `http://www.w3.org/WAI/eval/Overview.html`
8. Connor, J.O. Choosing an Accessible CMS, (extracted on June 2011 from), `http://juicystudio.com/article/choosing-an-accessible-cms.php`
9. NCDAE Tips and Tools: Content Management Systems & Accessibility, (extracted on June 2011 from), `http://www.ncdae.org/tools/factsheets/cms.cfm`
10. López, J.M., Pascual, A., Granollers, A.: Engineering Accessibility in Web Content Management System Environments. In: Federrath, H. (ed.) Designing Privacy Enhancing Technologies. LNCS, vol. 2009, pp. 597–604. Springer, Heidelberg (2001)

Instructional Animations: More Complex to Learn from Than at First Sight?

Anna Wong[1], Nadine Marcus[1], and John Sweller[2]

[1] School of Computer Science and Engineering,
University of New South Wales, Sydney, Australia
{annawong,nadinem}@cse.unsw.edu.au
[2] School of Education,
University of New South Wales, Sydney, Australia
j.sweller@unsw.edu.au

Abstract. This paper presents a cognitively guided set of design guidelines for instructional animations based on a review of the existing literature. The guidelines are based around a cognitive load theory framework, which assumes learners' limited working memories must be considered for instructional animations to be effective. We propose six design guidelines: (1) animations are more beneficial for learners with higher levels of prior knowledge; (2) animations are more effective for certain knowledge domains; (3) segment animations in shorter sections; (4) give learners control; (5) signal or cue important information; (6) remove details and information from instructional animations that are not necessary for learning.

Keywords: Cognitive load theory, animation, multimedia instructional design.

1 Introduction

Instructional animations are commonly used in many different types of user interfaces, particularly in computer-based educational environments. Additionally, the increasingly common use of new UI technologies such as interactive whiteboards and multi-touch devices, combined with increasing affordability of high-bandwidth Internet and the popularity of video streaming sites, make animations increasingly accessible in the classroom. Despite the appeal of animations, the factors influencing their effectiveness are not fully understood, often making appropriate implementation in educational settings difficult.

Instructional animations are a series of still frames that, when considered together, are perceived as motion, used for educational purposes. Existing research in animation has produced mixed results. Many studies show no intrinsic superiority of animation over static graphics [1], unless other design techniques to support learning were also introduced, such as learner control [2] or segmentation into shorter sections [3]. Other research shows animation to be highly effective for learning, but only when depicting human movement tasks [4]. The range of results indicates the effectiveness of animations is heavily influenced by a number of factors, making design guidelines even more critical to designers and educators alike.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 552–555, 2011.

Instructional animations are often difficult to learn from because they are *transient* [5]. Information appears on the screen and then disappears in a very short period of time, making it very difficult for information to be reviewed. Humans possess a working memory that can only store and process small amounts of information very briefly [6]. Our working memory limitations combined with the transience of animations mean that novel, complex information presented in an animated format can be challenging for learners to process. Transience inherent in most instructional animations means learners need to simultaneously remember and process both previously presented information as well as currently presented information to understand the learning material. However, previously learned information may have already been lost from working memory before the current information has been processed. In contrast, static graphics can be revisited on demand in a way that is more difficult to replicate using animations. The ability to revisit information in static graphics means previously presented information does not need to be held in working memory, thus reducing or potentially even eliminating transience and making materials easier to understand.

Although other design guidelines for instructional animations do exist, the pace at which technology evolves means they quickly become outdated (e.g., [7]), or are highly theoretical (e.g., [8]). Recent cognitive load related research can explain when and how animations can be effective.

1.1 Cognitive Load Theory

Cognitive load theory (CLT) [9] is a framework of research-based instructional design principles based on the characteristics and relationships between the structures within our human cognitive architecture, which persist regardless of age.

CLT is based around the notion of a working memory used to process current information, that is very limited in terms of both the quantity of novel information it can process and the duration such information can be stored [6]. In contrast, long-term memory is effectively infinite in terms of the quantity and duration it can store knowledge. Long-term memory is able to store such large amounts of information efficiently by organizing them into schemas, cognitive structures that help organize information according to how it will be used. Schemas brought in to working memory from long-term memory are treated as a single item in working memory, hence reducing working memory load. Central to CLT is the concept of cognitive load [9], the amount of mental effort exerted during learning or performing a task. There are three major types of cognitive load:

Intrinsic cognitive load is the cognitive load inherent in the learning materials. Intrinsic cognitive load cannot be altered without changing the meaningfulness of the content. It is determined primarily by the number of elements of information that needs to be considered simultaneously in working memory.

Extraneous cognitive load is the cognitive load arising from instructional design factors. This form of cognitive load can be changed through effective instructional design. Most classical cognitive load theory research has been concerned with lowering extraneous load.

Germane cognitive load is the load used to create schemas and automate them.

CLT has been used successfully over the last 20 years to guide instructional designers. More recently, it has been successfully used to inform multimodal interface design [10]. The guidelines generated by CLT can help inform both design and use of instructional animations.

2 Design and Teaching Implications

As mentioned previously, a primary characteristic of animation is its transience, which can overload learners' limited working memory. As such, design techniques associated with animations primarily aim to decrease transience. The CLT research to date suggests the following design and teaching recommendations:

Instructional animations are better for learners with more prior knowledge. None of the design principles below can be considered in isolation of learners' level of prior knowledge. Prior knowledge stored in long-term memory in the form of schemas, determines how we make sense of information. Design of instructional materials must account for learners' prior knowledge. Animations are more beneficial for learners who already have some prior knowledge in an area [11]. Prior knowledge reduces transience, hence reducing or even eliminating working memory limitations because limitations only occur when dealing with novel information. Low prior knowledge learners benefit less from animation, as they do not have the schemas to attenuate transient information in animations.

Instructional animations are more useful for some knowledge domains over others. There is emerging evidence that animations depicting some specific domains, such as human movement, are more efficiently taught using animations [4]. This may be because animations depicting motor skills tap into our innate ability to learn through observing other people, thus reducing the load on working memory.

Segment instructional animations into shorter sections. Segmenting animations into smaller sections results in better learning compared to non-segmented animations, especially if the segments are combined with simple forms of learner control, for example letting learners move between sections at will [2]. Segmenting animations reduces the amount of transient information to be processed by working memory.

Give learners some control over the animation. Simple forms of learner control, such as pacing, can improve the effectiveness of animations [2]. However, it should not be relied upon in the absence of appropriate instructional design, as students may not use it appropriately, or at all [5]. Learner control should also be designed to accommodate prior knowledge - learners with low prior knowledge in the knowledge domain to be learnt and the software environment may benefit more by having access to simpler forms of learner control until they develop expertise, when more complex forms of learner control can be introduced.

Signal or cue to focus attention on what's important. Signaling or cueing important parts of an animation can help focus the learner's attention on the critical parts [3], decreasing working memory load imposed by visual search. However, cues need to be designed carefully to not distract learners' attention from the information to be learnt.

Remove details and information from instructional animations that are not necessary for learning. Removing unnecessary information and details decreases extraneous cognitive load, making it easier for learners to focus attention where it is required [12]. Content and details chosen for inclusion in instructional animations should not distract from the focus of the learning task. Learners with high prior knowledge may be able to handle more detailed animations than novice learners.

3 Conclusions

Although new technologies show immense promise, it is important to understand learners have limited working memories. Instructional animations must be used and designed with these limitations in mind. The animation related guidelines listed are by no means complete and the use of animations still an active area of research. Rapid changes in technology mean that these guidelines will also continue to evolve. These guidelines also do not negate the need for appropriate instructional design of other learning materials. Nonetheless, we believe CLT and these guidelines can provide a useful learner-centered framework to help guide designers and educators alike.

References

1. Tversky, B., Morrison, J.B., Betrancourt, M.: Animation: can it facilitate? Int. J. Hum-Comput. St. 57, 247–262 (2002)
2. Mayer, R.E., Chandler, P.: When learning is just a click away: Does simple user interaction foster deeper understanding of multimedia messages? J. Educ. Psychol. 93, 390–397 (2001)
3. Moreno, R.: Optimising learning from animations by minimizing cognitive load: Cognitive and affective consequences of signaling and segmentation methods. Appl. Cognitive Psych. 21, 765–781 (2007)
4. Wong, A., Marcus, N., Ayres, P., Smith, L., Cooper, G.A., Paas, F., Sweller, J.: Instructional animations can be superior to statics when learning human motor skills. Comput. Hum. Behav. 25, 339–347 (2009)
5. Hasler, B.S., Kersten, B., Sweller, J.: Learner control, cognitive load and instructional animation. Appl. Cognitive Psych. 21, 713–729 (2007)
6. Cowan, N.: The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behav. Brain Sci. 24, 87–185 (2000)
7. Nielsen, J.: Guidelines for multimedia on the web, `http://www.useit.com/alertbox/9512.html`
8. Ayres, P., Paas, F.: Can the cognitive load approach make instructional animations more effective? Appl. Cognitive Psych. 21, 811–820 (2007)
9. Sweller, J.: Instructional Design in Technical Areas. ACER Press, Camberwell (1999)
10. Oviatt, S.: Human-centered design meets cognitive load theory: Designing interfaces that help people think. In: Proceeding of the 16th ACM International Conference on Multimedia, pp. 871–880. ACM Press, New York (2006)
11. Kalyuga, S.: Relative effectiveness of animated and static diagrams: An effect of learner prior knowledge. Comput. Hum. Behav. 24, 852–861 (2008)
12. Mayer, R.E., Heiser, J., Lonn, S.: Cognitive constraints on multimedia learning: When presenting more material results in less understanding. J. Educ. Psychol. 93, 187–198 (2001)

Integrating Feedback into Wearable Controls

Cátia Sousa and Ian Oakley

Madeira Interactive Technologies Institute,
University of Madeira, Caminho da Penteada, Funchal, Portugal
csfs.ferreira@gmail.com, ian@uma.pt

Abstract. Wearable computing is a highly specialized application domain requiring the development of novel interaction technologies. This is due not only to the distracted and eyes busy scenarios that such systems target, but also due to the necessity of implementing highly wearable systems that take advantage of the affordances of cloth and clothing. This paper presents the design and development of three novel wearable input devices based on commonplace elements of clothing: zippers, cords strung with beads and fabric patches. These devices implement different forms of input (respectively, linear continuous, linear discrete and tagging) and are notable in that they combine input with output in the form of persistent physical feedback – each device is designed to physically and visually resemble the digital information it controls. This paper argues that this approach is novel in the domain of wearable computing and has the potential to significantly improve usability.

Keywords: Smart textiles, wearable technology, craft materials, sensors.

1 Introduction

Integrating computation into human clothing is an intrinsically appealing idea. Wearable computers can aspire to be as ubiquitous as other practical tools built into our clothing (such as pockets) while providing access to sophisticated digital services, remaining continuously available for use and avoiding the burden of encumbering users with a discrete device such as a mobile phone. Careful design can fit wearable computers to our bodies for hands-free and eyes-free interaction in situations such as medical or sporting scenarios [1] and via emerging paradigms such as gestural interaction [2]. The flexible nature of cloth and clothing also affords novel opportunities for input, such as pulling or stretching cord-like appendages [3] or pinching different volumes of fabric [4]. These recent efforts to develop novel interaction techniques highlight the importance of designing wearable technologies and sensors that leverage the natural properties of fabric and clothing.

The work in this paper addresses this space. Its contribution is the design and implementation of three novel input techniques for wearable computers based on typical elements of clothing: a zipper, a bead on a cord and a system of removable fabric patches. These techniques are notable in that they each enable a different quality of input (respectively, linear continuous, linear discrete and tagging or identification) and are combined with persistent and visible physical alterations to clothing. The zipper is moved to continuously adjust a variable, the bead threaded

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 556–559, 2011.

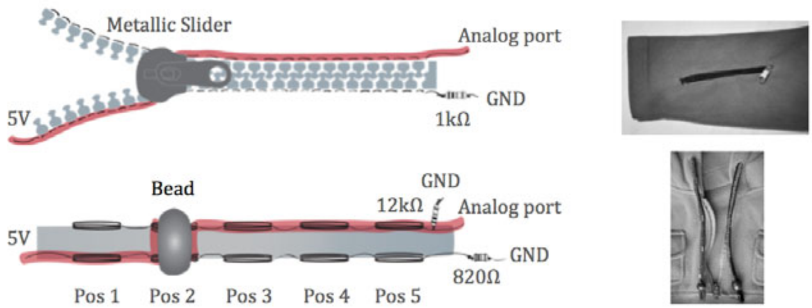


Fig. 1. Schematics for Zip and Toggle sensors (left). Sensors integrated into clothing (right).

along the cord to select specific modes and the patches added or removed to signify their presence or absence. Due to the complexity of embedding displays in cloth, such persistent feedback and state information is a rarity in wearable input techniques and this paper argues that its combination of rich input (when contrasted to commonly available binary pushbuttons [5]) coupled with visible feedback and output is a compelling new design space for research in interaction techniques for wearable computing. The introduction of this kind of coupled input/output will not only make wearable devices easier to use, but also offer designers and users new opportunities for making social and aesthetic statements.

2 Zip and Toggle Sensors

The first and second prototypes, based respectively on a *zipper* and a *toggle* (a cord strung with a bead) functioned as simple variable resistors. The zipper sensor was designed for continuous input while the toggle sensor allowed discrete, stepped input. Both devices incorporated two seams of 117/17 2-ply conductive thread, one sewn into each side. In both sensors, one of these lines connected a 5V signal to ground (via an appropriate resistor), while the other was wired to an input port on an Arduino microcontroller. In the zip sensor, the conductive thread was strung along the surface of the zip material for its entire length, while the cord sensor featured a design in which the thread was alternately exposed and hidden beneath the non-conductive material of the cord at intervals of approximately one centimeter. The devices also integrated a conductive movable element in contact with both lines of thread: the metal slider of the zip and the cord's bead, the inner surface of which had been lined with copper tape. By adjusting the position of these physical devices, the length of thread making up the circuit, and thereby its resistance and the voltage delivered to the Arduino's input port, varied systemically and measurably. Pictorials schematics for these devices, and examples sewn into clothing, can be seen in Figure 1.

These devices have a number of interesting properties. They are both familiar to users, have minimal costs and are highly wearable - the main component is two seams of conductive thread. They also intrinsically provide two forms of feedback: the physical texture felt when adjusting the moveable element and its persistent, visually

A

B

C

Fig. 2. Patch sensor A) front of final tag B) inner structure of tag showing rare-earth magnets at each corner C) jacket sleeve with conductive fabric; two tags are attached and two tags show circuitry on the back of the tags

observable position. The physical state of the device mirrors the state of the sensed digital data. They can also both be integrated into the functional fastenings of an item of clothing to measure aspects of the garment state – for instance whether or not a jacket or pocket is open or closed. Alternatively, they can be sewn into a non-functional zone to serve as a control device capable of manipulating a variable such as the volume in wearable media player (continuous input with the zipper) or the ring tone profile selected on a mobile phone (discrete input with the cord and bead). We believe these devices provide general-purpose functionality for wearable applications.

3 Patch Sensor

The third sensor was based on a series of small tokens, or *patches*, which could be added to or removed from a piece of clothing capable of both sensing their presence and distinguishing between them. It employed several technologies to achieve this functionality. At the most basic level, the tokens were built from 45mm by 45mm squares of 4mm Perspex. Small rare-earth magnets were attached to each corner of one side (the bottom) and the combined units encased in an outer layer of cloth. The surface of the bottom side was then augmented with a specific pattern of solder and copper tape while the top was used for distinctive graphical logos. Figure 2 shows a stripped down token and also examples of the top and bottom of complete units.

Figure 2 also shows the design of the apparatus that can sense and identify the tokens. It shows four "slots" in which tokens can be placed, two of which are occupied. Each slot features three rows of conductive fabric to which the patterns on the tokens connect. The bottom row is a 5V power rail shared by all the slots. The center row of each slot connects to a different Arduino input port and voltage on these channels is used to indicate whether a token is present or absent. The top row featuring the four smallest fabric electrodes is used to identify individual tokens – each also connects to a unique Arduino port. By connecting to different configurations of these electrodes 16 different binary patterns can be produced. Magnets positioned under the surface of the fabric ensure correct positioning and firm attachment of the tokens to the sensor unit.

Although this device has significant limitations, not least in the limited number of tokens it can support (up to fifteen different tags), it does provide interesting functionality. The tokens attach firmly to the clothing, but are easy to add and remove; the sensor (and tokens) are composed of simple, flexible, lightweight and washable components; the use of multiple slots supports ordering; the tokens can be associated with visual content (such as the musical content shown in Figure 2) to match their digital identities. This final quality is particularly interesting as the visual elements not only provide information about system state, but can also be used to represent aspects of a user's wearable computer in a public form – perhaps as statements about brand, fashion or identity.

4 Discussion and Conclusions

This paper has presented three novel wearable sensors: linear continuous input in a zip, discrete continuous input in a cord and bead and object identification via a series of removal patches. The range of devices described achieves one of the key goals of this work: to explore input of different qualities and expressiveness. The second focus was on how wearable input could be made visible and persistent to human users and observers. This has also been amply achieved: each of the three input devices directly senses information that is equally visible to a human – each device looks like the variable it represents. The sensors are also inexpensive, wearable and designed to integrate well into clothing by adapting existing accessories and styles.

There are many avenues for future work on this topic, including: functional evaluations of the robustness of the input mechanisms to use in real clothing; improving and developing the feedback present in the devices, for example by integrating visual designs on the clothing, or by incorporating physical cues; user studies capturing both qualitative opinions and quantitative performance with the devices. In conclusion, three novel wearable sensors that look like the data they represent (and control), were presented. This paper argues that such feedback and transparency is an incongruous rarity in wearable input and that the prototypes effectively illustrate how it can be achieved. In this way, they point the way for further work exploring this topic.

References

1. Bächlin, M., Förster, K., Tröster, G.: SwimMaster: a wearable assistant for swimmer. In: Proceedings of Ubicomp 2009, Orlando, pp. 215–224 (2009)
2. Rekimoto, J.: GestureWrist and GesturePad: unobtrusive wearable interaction devices. In: Proceedings of the Fifth International Symposium on Wearable Computers, Zurich, pp. 21–27 (2002)
3. Schwarz, J., Harrison, C., Hudson, S., Mankoff, J.: Cord input: an intuitive, high-accuracy, multi-degree-of-freedom input method for mobile devices. In: CHI 2010, pp. 1657–1660. ACM, New York (2010)
4. Karrer, T., Wittenhagen, M., Heller, F., Borchers, J.: Pinstripe: eyes-free continuous input anywhere on interactive clothing. In: Adjunct Proceedings of UIST 2010, pp. 429–430. ACM, New York (2010)
5. Fibretronic Linear Keypad, `http://www.fibretronic.com/connectedwear/KP25SH`

Intelligent Playgrounds: Measuring and Affecting Social Inclusion in Schools

Olga Lyra[1], Evangelos Karapanos[2], and Vassilis Kostakos[2]

[1] Centro de Competência de Artes e Humanidades
[2] Madeira Interactive Technologies Institute,
Universidade da Madeira
lyra.olga@gmail.com, {ekarapanos,vk}@uma.pt

Abstract. Equal access to education has recently been declared by the United Nations a basic human right [14]. However, despite the substantial attention given to inclusive education in recent years, researchers have criticized a lack of empirical evidence on how exclusion is manifested in student behavior. Recent development in sensor technology and social network analysis techniques can provide a new perspective to the impact of educational practices through the measurement of students' social interactions. In this paper we outline our research agenda that aims at a) measuring the current stand as well as the impact of inclusive educational interventions using sensor technology and sociometric analysis, and b) challenging pupils' perceptions of diversity with the aim of eliminating discriminatory behaviors in primary schools using persuasive games.

Keywords: Inclusive education, social networks, persuasive games.

1 Introduction

Inclusive Education represents an approach to education that seeks 'Education for All' through developing schools that respond successfully to the diversity of all learners and their different needs [15]. Equal access to education has recently been declared by the United Nations a basic human right. As of September 2009, most EU countries have signed the Ratification of the UN Convention whereby access to an inclusive educational system has become a legal right [14].

Schools have access to a number of operationalized techniques and methods for promoting inclusive school development, such as the Index for Inclusion [1], a tool being used in more than 40 countries. Index for Inclusion is designed to help schools in a) assessing their current stand with regard to inclusiveness, b) identifying the key topics of concern for the whole school community, and c) implementing concrete actions after reaching consensus among all members of the school community.

However, despite the substantial attention given to this educational approach in recent years, researchers have criticized a lack of empirical evidence on how exclusion is manifested in student behavior as well as the long-term effect of existing methodologies that promote inclusive school development [11]. Even when studies have attempted to capture how exclusion is manifested in students' social interactions

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 560–563, 2011.

within the class and during play time, their focus was limited to a number of school cases as well as particular dimensions of diversity, thus leading to an uncertainty of how such results may generalize to the larger population and how educational exclusion is manifested at large [6].

Recent development in sensor technology [4] and social network analysis techniques [10] can provide a new perspective to the impact of educational practices through the measurement of students' social interactions. For instance, proximity measurement during playtime capturing pupils' social interactions may be correlated with metrics of different dimensions of diversity such as ethnic origin, religion and native language, parents' socio-economic and educational background, forms of special needs and educational performance,. These in turn may affect a student's ability to participate in equal terms to the educational community. Capturing these metrics can possibly generate the ability to measure inclusiveness at large, through sampling a wide number of schools and over extended periods of time.

Secondly, grounded on data about pupils' social interactions and using persuasive techniques (Fogg, 2002), technology may be used in challenging pupils' perceptions of diversity with the aim of eliminating discriminatory behaviors in primary schools. In the remainder of the paper we outline our research agenda.

2 Measuring the Inclusiveness of School Communities Using Sensor Technology and Sociometric Analysis

Sociometric analysis has been used for describing communication patterns in organizations [4], inquiring into urban mobility [8], modeling knowledge propagation and other domains [2], leading to a new field of computational social science [9] motivated by our increasing ability to capture social phenomena at large. To our knowledge, however, such techniques have not yet been used to capture the development of pupils in schools. Using sensor technology to capture and quantify schools' inclusiveness provides the ability to generalize our findings as measurements can take place at a wide number of schools and through an extended period of time.

This objective will result in a palette of metrics that capture the inclusiveness of a school community based on pupils' social interactions. The metrics will take into account pupils' proximity data during school activities, physical activity (through accelerometer) and verbal activity (through microphone capture). In addition, dimensions of diversity such as ethnic origin, parents' socio-economic and educational background, forms special needs and others will be captured. These metrics will subsequently be used to assess the impact of the educational and technological interventions in the next two goals.

3 Assessing the Impact of Inclusive School Development Interventions

The impact of educational interventions, and specifically "Index for Inclusion" projects, will be assessed in terms of changes in pupils' social interactions, using sensor network technology. Index for inclusion projects will be executed in a number of schools and measurements of pupil's social interactions will be taken throughout the intervention. Pair-wise proximity will be the key measured variable but other variables such as physical activity (through accelerometer) and verbal activity (through microphone) will be captured. A combination of these measures will be used to express social activity, and will then be correlated with: a) Metrics of different dimensions of diversity that might affect a student's ability to participate in equal terms to the educational community such as ethnic origin, religion and native language, parents' socio-economic and educational background, forms special needs and educational performance, and b) stakeholders' perceptions of the inclusiveness of a school, and data elicited in the course of Index for Inclusion projects.

The analysis will look to develop models of how students interact with others, and will attempt to assess whether the social structure amongst students impacts their performance in school. This analysis will control for the various environmental variables that change from school to school.

4 Persuasive Games That Challenge Children's Perceptions of Diversity and Demotivate Discriminatory Behaviors

Construal theory [13] suggests that (perceived) social similarity, being it in the form of attitudes, physical characteristics etc., influences not only the available information, but also the cognitive process involved when individuals evaluate other's actions, with more distant individuals to be judged on more abstract attributes. It has been found that when individuals evaluate their own, and those of similar others, behavior, they emphasize the role of concrete situation factors that operate at the moment of action ("I stepped on your toe because the bus was crowded"), whereas when judging the behavior of more distant individuals, they emphasize the role of stable, general dispositional properties of the actor ("he stepped on my toe because he is clumsy") [13].

"Daily habits" is a persuasive game that attempts to increase of interpersonal similarity through capturing and communicating habits that are shared across different children that appear limited interaction in school. Habits will be sensed through the mobile phones and may include aspects of a child's daily life such as: wake-up time, sleeping time, or the amount, temporal and spatial range of physical activity.

However, motivating social behaviors brings significant complications, for instance, through group polarization [7], a phenomenon where people adopt more extreme response to deliberate acts when being in groups than when being alone.

Through minimalistic interventions we will attempt to measure the impact of two motivational techniques: *reciprocity* and *operant conditioning*. Reciprocity has been found to have positive effect in encouraging prosocial behaviors in primary schools [3]. It may be operationalized through the identification of positive changes in a pupil's social behavior (e.g. his or her centrality in the social network) and the communication of those to other pupils in the class. Operant conditioning [12] uses two primary tools: *reinforcement* and *punishment*. These may be operationalized in the form of a game where positive daily changes in a child's network centrality adds points to the child's profile while negative changes remove points.

5 Conclusion

This paper presents a double-pronged research agenda aiming first to measure the status quo and impact of inclusive educational interventions using sensor technology and sociometric analysis. In addition, it aims to challenge pupils' perceptions of diversity with the aim of eliminating discriminatory behaviors in primary schools using persuasive games.

References

1. Booth, T., Ainscow, M.: Index for Inclusion. In: Developing Learning and Participation in Schools, 2nd edn. Centre for Studies on Inclusive Education, Bristol (2002)
2. Christensen, C., Albert, I., Grenfell, B., Albert, R.: Disease dynamics in a dynamic social network. Physica A: Statistical Mechanics and its Applications 389(13), 2663–2674 (2010) ISSN 0378-4371
3. Covington, M.V.: Goal theory, motivation, and school achievement: an integrative review. Annual Review of Psychology 51, 171–200 (2000)
4. Eagle, N., Pentland, A.: Reality Mining: Sensing Complex Social Systems. J. Personal and Ubiquitous Computing (2005)
5. Fogg, B.J.: Persuasive Technology: Using Computers to Change what We Think and do. In: Grudin, J., Nielsen, J., Card, S. (eds.) 1st edn., Science & Technology Books (2002)
6. Fryxell, D., Kennedy, C.: Placement along the continuum of services and its impact on students' social relationships. Journal of the Association for Persons with Severe Handicaps 20, 259–269 (1995)
7. Isenberg, D.J.: Group polarization: A critical review and meta-analysis. Journal of Personality and Social Psychology 50(6), 1141–1151 (1986)
8. Kostakos, V., O'Neill, E., Penn, A., Roussos, G., Papadongonas, D.: Brief encounters: sensing, modelling and visualizing urban mobility and copresence networks. ACM Transactions on Computer Human Interaction 17(1), 1–38 (2010)
9. Lazer, D., Pentland, A., Adamic, L., Aral, S., Barabasi, A.-L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., Van Alstyne, M.: Computational Social Science. Science 323, 721–723 (2009)
10. Macindoe, O., Richards, W.: Graph Comparison using Fine Structure Analysis. In: Proc. IEEE SocCom 2010, Minneapolis, USA (2010)
11. Simpson, R.L.: Inclusion of Students with Behavior Disorders in General Education Settings: Research and Measurement Issues. Behavioral Disorders 30(1), 19–31 (2004)
12. Skinner, B.F.: The Technology of Teaching. Appleton-Century-Crofts, New York (1968)
13. Trope, Y., Liberman, N.: Construal-Level Theory of Psychological Distance. Psychological Review 117(2), 440–463 (2010); 463a, ISSN 0033-295X
14. UN: Convention on the Rights of Persons with Disabilities (2006)
15. UNESCO (1994): The Salamanca Statement and Framework for Action on Special Needs Education. Salamanca, Spain, June 7-10 (1994)

It Does Not Fitts My Data! Analysing Large Amounts of Mobile Touch Data

Niels Henze and Susanne Boll

University of Oldenburg,
Escherweg 2, 26121 Oldenburg, Germany
{niels.henze,susanne.boll}@uni-oldenburg.de

Abstract. Touchscreens are the dominant input device for smartphones and learning about smartphone users' touch behaviour became even more important. We developed a game for Android phones to collect a truly large amount of touch data from diverse devices and players. A part of the game is designed as what we expected to be a Fitts' law task. By publishing the game in the Android Market we collected 5,359,650 micro tasks from 63,154 installations of the game. Using Fitts' law to find a model for these tasks we found a very weak correlation and an implausible high index of performance across different devices. Further analysis shows a similar correlation between time and distance as with Fitts' law but only a very weak correlation with the targets' width.

Keywords: Fitts' law, mobile phone, touch screen, app store, large-scale.

1 Background

Fitts' law [2] is a model of human movement that predicts the time required to rapidly move to a target. It describes the correlation between the distance to a target and it's width with the time to reach the target. [3] describes Fitts' law as "one of the most robust, highly cited, and widely adopted models to emerge from experimental psychology" [3]. The index of performance (IP), that is determined using Fitts' law, enables to compare different input devices and is even proposed by ISO 9241-9 [1] for this purpose. Fitts' law has been applied to a wide range of devices, including mouse, touchpads, and trackballs [3]. The IP was used to compare these devices and we assume that the IP can also be used to assess the quality of smart phones' touch screens.

To collect data from a number of different devices, in a natural usage context, and a sufficient number of participants we developed a game and published it in the Android market. The game requires touching targets and records the players' behaviour. A part of the game is a task that requires to sequentially tapping simultaneously presented circles. Using data from 63,154 installations of the game and considering tapping each but the first of the simultaneously presented circles as a task we collected 5,359,650 tasks.

2 Collecting Touch Data in the Large

The gameplay of Hit It! is inspired by the task used by Park et al. [4]. Circles are displayed on the screen (see Figure 1) and the player has to touch these targets. Each level consists of multiple micro levels. Different target sizes are used in each level. In

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 564–567, 2011.

most micro levels, one circle is presented to the player. The player advances directly to the next micro level as soon as the target is hit. If a target has not successfully been hit in a certain time frame it is counted as a miss. Every fifth micro levels consist of multiple simultaneously presented targets. As soon as a target is successfully hit it disappears. The player must hit all targets to advance to the next micro level. To make a game out of the two basic tasks the player must complete a micro level in a certain timeframe. The time is reduced from micro level to micro level while the player proceeds through a level. The player receives a penalty point if a target has not been hit and the game is lost when the player collected three penalty points. The faster players hit the targets the more scores they get.

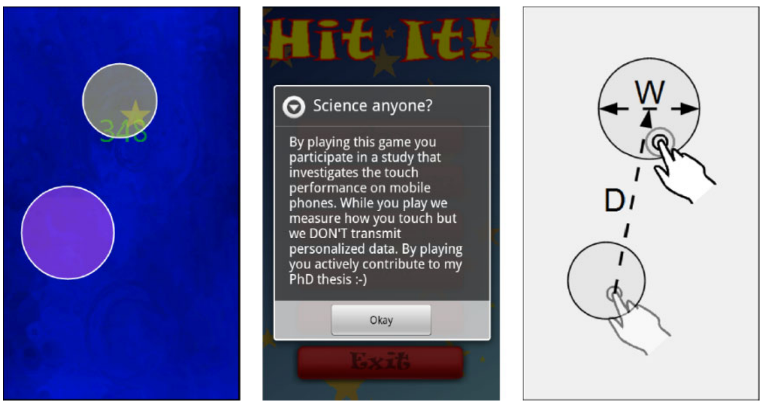


Fig. 1. In-game screenshots of the game (left), information presented to the player when the game is started (centre), and how W & D are derived for Fitts' law (right)

We do not collect data that allows identifying individual players or installations. We still decided to clearly inform players about the fact that data is collected in order to act ethically. A modal popup (see Figure 1) shown when the game is started for the first time tells the players that data is collected for research purpose while the game is played. Data about the used device and the performance of a player is collected by transmitted the data to our server. A unique identifier for each installation is derived by hashing the device's “Android ID”. We also collect the user's locale (e.g. "en_GB") and the device's type (e.g. "GT-I9000" for Samsung Galaxy S). Most importantly, we record the position and size of the targets for each micro level. We also record the position of each touch event and the time elapsed since the start of the micro level. We published Hit It! in the Android Market on October 31, 2010. Till April 04, 2011 the game was installed 192,250 times[1].

3 Applying Fitts' Law

When developing the gameplay we designed two tasks. The first task is just tapping a single target but the second task was designed to be a Fitts' law task. In the following

[1] Number of installs according to Google's Developer Console.

we analyze the data from the second task using Fitts' law and investigate the correlation of movement time with the target's distance and width afterwards.

3.1 Correlation and Throughput

In order to apply Fitts' law we use the Shannon formulation $MT = a + b \cdot log_2\left(1 + \frac{D}{W}\right)$ proposed by [3]. We only consider micro levels where multiple circles are presented simultaneously, the device is held in an upright posture (according to the phone's accelerometer), and the level's error rate is below 4%. When the player hits the first circle the touched position and the current time is used as the start for the first task (see Figure 1). MT is the time that elapsed until the second circle is hit. D is the distance from the position of the first tap to the centre of the second circle and W is the diameter of the second circle. Analogical we proceeded with the other targets using the touched position that hit one circle and the according subsequent circle. Thereby multiple tasks are derived from a single micro level.

With the described approach we determined 5,359,650 tasks from 63,154 installations. Using linear regression we determined a (intercept) and b (slope) and found that a=.200 and b=.040. The correlation of the data with the resulting formula is r=.14. The index of performance ($IP = \frac{1}{b}$) is IP=25.01. As the determined IP is implausibly high (highest IP reported by [3] is 13.7 for an eye tracker) and the correlation is very weak ([3] reports correlations between .70 and .99) we selected the considered data more carefully. Using only data provided by 4,873 installations running on a Samsung Galaxy S (292,515 tasks), the most common device in our dataset, results in a=.207, b=.029, and r=.13. With 34.57 the IP is still implausibly high. Considering only tasks from one particular level (992 installations, 5,328 tasks) results in a=.176, b=.060, r=.33, and IP=16.55.

One reason for the unsatisfying correlation might be the small screen size of the Samsung Galaxy S (4.0in diagonal size) that does not require moving the hand but only one finger. Therefore, we repeated the analysis for the Samsung Galaxy Tab (352 installations, 2,134 tasks) that has a screen with 7.0in. Again only tasks from one level are considered. Using linear regression we determined intercept and slope and found a=.165 and b=.063. The correlation of the resulting formula is r=.33 and the index of performance is IP=15.80. Regardless of the way we constrain the dataset, the IP is unlikely high and the correlation is weak for all tested devices. While the high IP only suggest that the used task might be different from what is commonly used as Fitts' law task, the rather low correlation shows that Fitts' law is not a good model for the used tasks.

3.2 Correlation with Targets' Width and Distance

In order to investigate why Fitts' law cannot be used as a model for our data we analyzed the data in more detail. We tested the correlation with the target's width and the distance to the target independently. For a true Fitts' law task one would expect that there is at least a weak correlation with both aspects that are composed in Fitts' law. As the measurement unit is not crossed out when looking at width and distance separately we have to treat each device independently.

Using again the data from one level provided by the Samsung Galaxy S (992 installations, 5,328 tasks), the correlation between time and distance is r=.33. Compared with the correlation using Fitts' law (also r=.33) the difference is very small. This suggests that most of the variance that can be explained by the correlation with ID can also be explained by the correlation with the distance to the target. Determine the Pearson correlation between time and width using the same dataset results in r=.11. Testing different functions (e.g. $\log_2(W)$, $1/W$, $\log_2(1/W)$) does not reveal a higher correlation.

4 Discussion and Future Work

Analyzing the data that resulted from players that touch a sequence of circles we found only a weak correlation when applying Fitts' law as a model for the data. Furthermore, the IP is implausible high. Further looking at the data we found a similar linear correlation between time and distance as if using Fitts' law. In contrast we found a much smaller correlation between time and the width of the target. Our results suggest that, for our task, the time required to hit a target partially depends on the distance to the target. A target's width, however, has a much smaller effect on the time required to hit a target.

The results presented in this paper are only a fraction of what can be investigated using the collected data. Further analysis showed that touch positions are systematically skewed and a compensation function that shifts the users' touches can reduce the error rate [5]. An area we consider interesting for future work is investigating touch sequences. How a target is hit might not only depend on the target's size and location but also on preceding touch contacts and the elapsed time. Applying such findings might result in a function that shifts the touch events to improve the users' touch precision that in generally applicable. We would like to share the data with other researchers to enable further analysis[2].

References

1. Douglas, S., Kirkpatrick, A., MacKenzie, I.: Testing pointing device performance and user assessment with the ISO 9241, Part 9 standard. In: Proc. CHI (1999)
2. Fitts, P.M.: The information capacity of the human motor system in controlling the amplitude of movement. Journal of Experimental Psychology 47 (1954)
3. MacKenzie, I.: Fitts' law as a research and design tool in human-computer interaction. Human-computer Interaction 7(1), 91–139 (1992)
4. Park, Y., Han, S., Park, J., Cho, Y.: Touch key design for target selection on a mobile phone. In: Proc. MobileHCI (2008)
5. Henze, N., Rukzio, E., Boll, S.: 100,000,000 Taps: Analysis and Improvement of Touch Performance in the Large. In: Proc. MobileHCI (2011)

[2] The complete dataset can be retrieved from `http://nhenze.net/?page_id=673`

Measuring Cognitive Workload with Low-Cost Electroencephalograph

Avi Knoll, Yang Wang, Fang Chen, Jie Xu,
Natalie Ruiz, Julien Epps, and Pega Zarjam

National ICT Australia
University of New South Wales
{yang.wang,fang.chen,natalie.ruiz,
julien.epps, pega.zarjam}@nicta.com.au
ajk@cse.unsw.edu.au, jie.jackxu@gmail.com

Abstract. Electroencephalography (EEG) is an important physiological index of cognitive workload. While previous research has employed high-end EEG devices, this work investigates the feasibility of measuring cognitive workload with a low-cost EEG system. In our experiment, EEG signals are recorded from subjects performing silent reading tasks under different difficulty levels. Experimental results demonstrate the effectiveness of cognitive workload evaluation even with low-cost EEG equipment.

Keywords: Cognitive workload, electroencephalography (EEG), physiological index.

1 Introduction

In recent years, research efforts have been geared towards measuring human mental states such as cognitive workload and task engagement. Cognitive workload refers to the amount of mental demand imposed by a particular task on a person [3]. Measuring cognitive workload is an important issue in various research and application areas of human-computer interaction, as it can be utilized to evaluate the efficacy of interfaces and build adaptive interaction systems. With the advance of modern sensing technologies, a variety of physiological measures have been developed for the assessment of cognitive workload. Among these techniques, electroencephalography (EEG) has become a popular physiological index that allows continuous monitoring of subjects' cognitive workload in a convenient way.

Previous research has demonstrated that EEG signals are sensitive to cognitive load changes in various tasks [1]. Gevins and Smith [5] demonstrated that spectral features of the theta and alpha frequency bands correlate with task difficulty levels in simulated flight tasks and n-back tests. Fitzgibbon et al. [4] also found that the power of gamma band could be augmented by various cognitive tasks. Berka et al. [2] employed discriminant function analysis on spectral features for monitoring cognitive workload and task engagement in different tasks including digit span, mental arithmetic, image learning and memory tests. Grimes et al. [6] and Zarjam et al. [9] investigated EEG based classification and evaluation of subjects' working memory

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 568–571, 2011.

load. A feature selection scheme based on mutual information was proposed in [6] to deal with physiological drift. EEG has also been used to monitor cognitive workload in various military tasks under complex environments [8].

Although EEG is a promising tool for continuous measurement of cognitive workload, most previous research has employed high-end EEG systems costing more than $15,000 (e.g. see www.biosemi.com), which limits their widespread usage in human-computer interfaces. On the other hand, low-cost (under $1000) EEG headsets have become accessible for HCI research in recent years [7]. This work takes an initial step in exploring the feasibility of cognitive workload evaluation using a low-cost multi-channel EEG system.

2 Experiment

Sixteen students and employees (16-46 years old, 4 females) were invited to perform silent reading tasks. Brain waves from each subject were recorded with a low-cost EEG device originally designed for gaming interfaces (Emotiv EPOC, a 14 channel 128Hz neuro-signal acquisition and processing wireless neuroheadset [10], see Figure 1). Channel names based on the International 10-20 locations are: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4. During the experiment, each subject was asked to silently read the text displayed on-screen, with a viewing distance of 70cm (see Figure 2). Similar to [9], different task difficulty levels were employed to manipulate cognitive workload during the experiment.

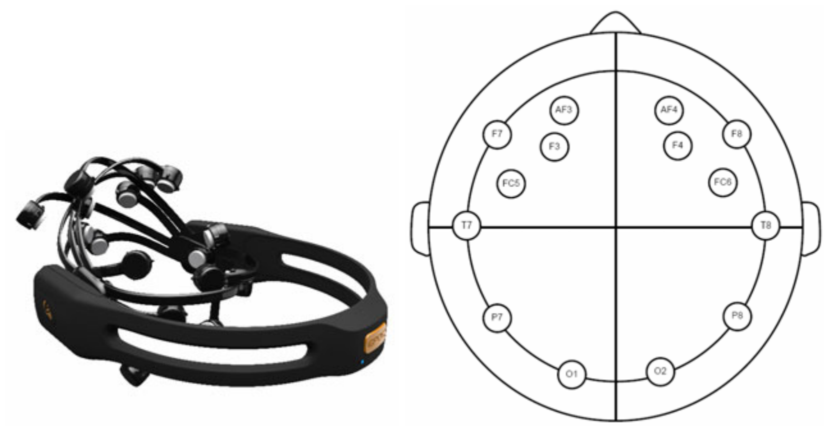


Fig. 1. A low-cost EEG device (Eomtiv EPOC neuroheadset [10]) and layout of EEG channels

There were three levels of task difficulty in total: low (level 1), medium (level 2) and high (level 3). For each difficulty level, 4 text pages were sequentially displayed on the screen, with each page appearing for 30 seconds. In the low level task, the subject was required to press the left mouse button when he encountered any 3 letter word during silent reading. In the medium level task, the subject was required to press the left or middle button respectively, each time he encountered either a 3 or a 4 letter word. Likewise, in the high level task, the subject was required to press the either the left, middle, or right button when he saw a 3, 4, or 5 letter word respectively. The task difficulty levels were administered in a randomized fashion. There was a 30 second resting period after the task for each difficulty level. One minute baseline data (with the subject looking at a blank screen) was recorded at both the beginning and the end of the whole experiment for each subject. The subject was asked to refrain from eye blinking and to stay as still as possible during the baseline period and task period. However, the subject was free to blink and move their head naturally during each rest period.

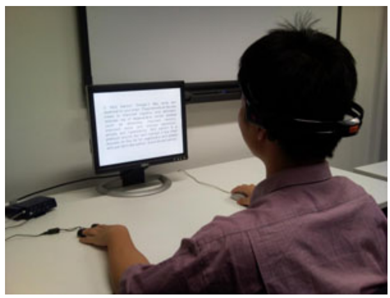


Fig. 2. Experiment setup

3 Analysis

The EEG signals were first divided into segments of 1.5 seconds in length. Statistical features including mean, variance, root mean square (RMS), spectral powers of theta (3-7 Hz), alpha (8-12 Hz), beta (13-29 Hz), and gamma (30+ Hz) frequency bands were then calculated for each data segment.

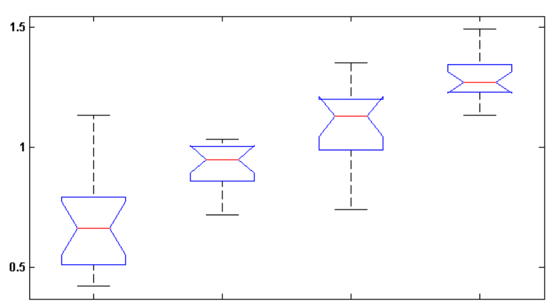


Fig. 3. Box plot of normalized RMS values (sample minimum, lower quartile, median, upper quartile, and maximum) from nodes F3 and F4 at different task difficulty levels

Among the features obtained from different EEG channels, RMS from nodes F3 and F4 exhibit significant correlation with task difficulty ($F > 38$, $p < 0.01$ in ANOVA test). This finding is consistent with previous research indicating that the brain frontal lobes play an important role in cognitive tasks associated with attention and mental effort [5]. Figure 3 plots the distribution of normalized RMS acquired from the two frontal channels at different workload levels for all the subjects. It can be seen that the feature value consistently increases when the task difficulty level is increased.

Moreover, the spectral power of gamma frequency band at nodes AF3 and AF4 shows a statistically significant difference between the baseline condition and task condition ($F > 28$, $p < 0.01$ in ANOVA test), which is consistent with previous study on gamma activation of EEG during cognitive tasks [4]. There is an increase in average gamma power with each rise in task difficulty. However, the difference between task levels is not statistically significant ($p > 0.05$).

4 Conclusion

This work investigates the feasibility of cognitive workload evaluation using a low-cost EEG system. It is demonstrated that cognitive workload could be effectively measured even with low-cost electroencephalograph. The experimental results are consistent with previous research on cognitive workload. We hope that this work will promote the application of EEG-based physiological measures in various HCI areas involving cognitive workload evaluation.

References

1. Antonenko, P., Paas, F., Grabner, R., van Gog, T.: Using electroencepahlography to measure cognitive load. Educational Psychology Review 22, 424–438 (2010)
2. Berka, C., Levendowski, D., Lumicao, M., Yau, A., Davis, G., Zivkovic, V., Olmstead, R., Tremoulet, P., Craven, P.: EEG correlates of task engagement and mental workload in vigilance, learning, and memory tasks. Aviation, Space, and Environmental Medicine 78, B231–B244 (2007)
3. Cain, B.: A review of the mental workload literature. Technical Report, Defence Research and Development Canada Toronto (2007)
4. Fitzgibbon, S., Pope, K., Mackenzie, L., Clark, C., Willoughby, J.: Cognitive tasks augment gamma EEG power. Clinical Neurophysiology 115, 1802–1809 (2004)
5. Gevins, A., Smith, M.: Neurophysiological measures of cognitive workload during human-computer interaction. Theoretical Issues in Ergonomic Science 4, 113–131 (2003)
6. Grimes, D., Tan, D., Hudson, S., Shenoy, P., Rao, R.: Feasibility and pragmatics of classifying working memory load with an electroencephalograph. In: Proc. ACM CHI Conf. Human Factors in Computing Systems, pp. 835–844 (2008)
7. Lee, J., Tan, D.: Using a low-cost electroencephalograph for task classification in HCI research. In: Proc. ACM Symposium on User Interface Software and Technology, pp. 81–90 (2006)
8. Wilson, G.: An analysis of mental workload in pilots during flight using multiple psychophysiological measures. International Journal of Aviation Psychology 12, 3–18 (2002)
9. Zarjam, P., Epps, J., Chen, F.: Evaluation of working memory load using EEG signals. In: Proc. APSIPA Annual Summit and Conference, pp. 715–719 (2010)
10. `http://www.emotiv.com/`

Model-Based Accessible User Interface Generation in Ubiquitous Environments

Raúl Miñón, Julio Abascal, Amaia Aizpurua, Idoia Cearreta, Borja Gamecho, and Nestor Garay

Informatika Fakultatea,
University of the Basque Country/Euskal Herriko Unibertsitatea,
Manuel Lardizabal 1, 20018 Donostia, Spain
{raul.minon,julio.abascal,amaia.aizpurua,idoia.cearreta, borja.gamecho,nestor.garay}@ehu.es

Abstract. This paper presents a system that automatically generates accessible interfaces tailored to the users' capabilities and needs in order to provide them with access to ubiquitous computing environments. The aim is to ensure that people with disabilities are able to use ubiquitous services provided by intelligent machines, such as ATMs and vending machines. The tailored interfaces are generated from a formal description specified by a User Interface Description Language, and based on user and context models represented by ontologies.

Keywords: Adapted user interfaces, accessibility, ubiquitous computing.

1 Introduction

People with physical, sensory or cognitive disabilities are frequently excluded from digital services since they cannot use interactive systems that have been designed ignoring human diversity. Due to the advances in mobile technology, ubiquitous services have flourished in diverse contexts. In a typical ubiquitous scenario a user equipped with a mobile device enters a ubiquitous environment, where a software layer -called middleware- enables access to the ubiquitous system, and the local services are offered to her or him. If the user selects one of the available services, a user interface is downloaded to the user device. In this way, the user can access the service through the user interface using her or his own handheld device.

In order to improve the autonomy and inclusion of people with disabilities, "intelligent" machines (such as information kiosks, ATMs, vending machines, home appliances, etc.) must offer ubiquitous services in an accessible way. This will enable users to conveniently access these services through their mobile device, which is adapted to their needs and preferences. Nevertheless, service providers usually offer a unique user interface for all users, which usually does not cater to every user's needs. Therefore, a user may experience accessibility problems when trying to access such services.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 572–575, 2011.

1.1 Related Work

Among the abundant works on adapted user interaction, only a few research works are devoted to adapting the content, presentation or the navigation scheme of the user interface to users with special needs. Nevertheless, some interesting works can be found in the literature [1, 2].

In the context of adaptive systems, the adaptations are generally applied to a previously created user interface. However, in the case of adaptive systems for ubiquitous environments, the adaptation process is not based on an existing user interface but on an abstract description of the functionality of the service [3]. This entails new challenges, such as determining the structure and organization of the elements that will comprise the final user interface. In addition, the accessibility of the interface can be considered before creating it. This approach avoids the need to find and eliminate the accessibility barriers that may exist in a previously designed user interface.

These two research lines (adaptive accessible user interfaces for people with special needs and automatic generation of user interfaces for ubiquitous environments) were combined in the Inredis[1] project. This project was intended to provide users with disabilities with access to ubiquitous services through their own mobile devices in an accessible manner. One of the outcomes of the project was the software component called *Interface Generator* (IG). This module was responsible for creating a final user interface adapted to each user's needs, enabling the use of the service without accessibility barriers.

This paper describes the approach that was followed in the project for automatically generating accessible user interfaces. The rest of the paper is organized as follows: in section two, the architecture of the Interface Generator is presented; the features of the user interfaces are explained in section three; finally, some conclusions are summarized in section four.

2 Architecture of the Interface Generator

The IG component follows a modular architecture, which is sufficiently flexible to accommodate new modules in the future. Each module of the IG performs a specific function, as explained below:

The **Orchestrator** module is in charge of organizing all the processes in the system. It plays a key role in the independence of the modules and it also enables the integration of new modules. The **Resource manager** is responsible for analyzing the resources offered by the provider of each service, and selecting the resource types (text, images, etc.) that are most appropriate for the user and her or his device. The **Constructor** creates a user interface, without any adaptations, from a User Interface Description Language. The **Selector** chooses the required adaptations by considering the users' capabilities, in order to provide them with an accessible user interface. Two

[1] `http://www.inredis.es/Default.aspx`

examples of adaptations are "enlarging text size" and "changing the contrast". The adaptations are implemented with EXtensible Stylesheet Language[2] (XSL) transformations and Cascading Style Sheets[3] (CSS) code fragments. The **Adapter** applies the previously selected adaptations into the interface created by the Constructor. In this way, an accessible interface that is also adapted to the users' needs is obtained. The **Data Injector** first checks whether a user interface already exists for the user, her or his device and the service. If so, the interface is downloaded to the user device instead of repeating the whole generation process.

In addition to the aforementioned modules, the IG makes use of a *Knowledge Base* (KB) which is composed of different ontologies for each of the required models: user, access device and context. The ontologies are queried by the IG during the interface generation process in order to retrieve information from the models.

Apart from these ontological models, which are focused on the user side, the IG employs user interface abstract models [4]. These models enable the automatic generation of interfaces regardless of the underlying technology (e.g. system platform or markup language). Starting from an abstract description of the functions provided by the service, the system is able to generate a functional user interface that allows users to access ubiquitous services and devices. In the particular case of Inredis, the User Interface Markup Language[4] (UIML) was chosen because it was the language that best suited the project requirements.

3 User Interfaces

The final user interfaces automatically generated by the IG are focused on certain user stereotypes. The system generates different interface alternatives for the same service depending on the person who is interacting. In order to demonstrate the viability of the system, three different web-based user interfaces were selected to be generated: a text-only interface, an iconic interface and a mixed interface that incorporates both textual and iconic elements.

The text-only user interface is divided into two main parts. In the first part the interaction commands are placed, whereas in the second part each command is described to facilitate the use and understanding of the service functions. This interface is aimed at blind people using screen readers, as well as users with mobile devices that do not support advanced resources.

All the elements in the iconic interface are represented as icons. The interaction elements for each function of the service are grouped into sections to allow users to better understand the interface. This version would be valid for people with cognitive disabilities, prelingual hearing-impairments, and illiteracy.

The mixed interface would be useful for users with mild sensory restrictions; for example, elderly people or users with little experience in using computing technology. The aim is to create a redundant interface that contains simple text to label the

[2] http://www.w3.org/TR/xsl/

[3] http://www.w3.org/Style/CSS/

[4] http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=uiml

interaction elements, accompanied by icons that emphasize the meaning of those tags. For this reason, every element in the interface has both textual and iconic resources to be grouped together.

All the generated user interfaces comply with XHTML Basic 1.1[5] syntax. To this end, the transformation and adaptation rules were implemented with XHTML syntax requirements. In addition, appropriate subsets of the Web Content Accessibility Guidelines[6] (WCAG) were considered in the generation rules, according to the type of interface and the target stereotype.

4 Conclusions

In this paper we have described the Interface Generator software component, which was one of the outcomes of the Inredis project. The prototypes developed demonstrate that the approach followed works well for the automated generation of adapted accessible user interfaces for different types of users in the context of ubiquitous environments. Although the interfaces generated are fully functional, further user testing is required in order to verify the validity and appropriateness of the interfaces for each type of user, as well as to check the accessibility and usability of the generated interfaces for each target user.

Acknowledgments. This research work has been partly funded by the Department of Education, Universities and Research of the Basque Government. In addition, A. Aizpurua, B. Gamecho and R. Miñón enjoy PhD scholarships from the Research Staff Training Programme of the Department of Education, Universities and Research of the Basque Government. We thank the Inredis project, which has been the foundation for this work.

References

1. Fink, J., Kobsa, A., Nill, A.: Adaptable and Adaptive Information Provision for All Users, Including Disabled and Elderly People. New Review of Hypermedia and Multimedia 4(1), 163–188 (1998)
2. Gajos, K.Z., Weld, D.S., Wobbrock, J.O.: Automatically generating personalized user interfaces with Supple. J. Artificial Intelligence 174(12-13), 910–950 (2010)
3. Hervás, R., Bravo, J.: Towards the ubiquitous visualization: Adaptive user-interfaces based on the Semantic Web. Interacting with Computers 23(1), 40–56 (2011)
4. Guerrero-Garcia, J., Gonzalez-Calleros, J.M., Vanderdonckt, J., Munoz-Arteaga, J.: A theoretical survey of user interface description languages: Preliminary results. In: The Latin American Web Congress (LA-WEB 2009), pp. 36–43. IEEE Computer Society, Washington, DC (2009)

[5] http://www.w3.org/TR/xhtml-basic/
[6] http://www.w3.org/TR/WCAG20/

Multiuser Augmented Reality System for Indoor Exhibitions

Jesus Gimeno, Ricardo Olanda,
Bibiana Martinez, and Fernando M. Sanchez

Instituto IRTIC,
Universidad de Valencia,
Calle Catedrático José Beltrán 2,
46980 Paterna, Valencia, España
{Jesus.Gimeno,Ricardo.Olanda,Bibiana.Martinez,
Fernando.M.Sanchez}@uv.es

Abstract. Over the last years, museums and galleries are looking for new ways to show exhibitions to visitors. For that purpose, new technologies like augmented reality are used. In this paper an augmented reality system for indoor exhibitions is presented. The system is formed by visualization screens that mix exhibition environment, visitors included, with multimedia and virtual 3D objects which visitors can manipulate naturally using a markers system. This system has been used in the exhibition "Valencia, tierra de comarcas: Diálogos con el patrimonio", which deals with a trip through the valencian cultural heritage.

Keywords: Augmented reality, interaction, augmented mirror.

1 Introduction and Goals

Over the last years different technologies have provided museums and galleries visitors with a new form to see their stored knowledge. One of these technologies is augmented reality (AR). AR is an interaction paradigm that aims to combine computer-generated information with the real world [1].

Through AR within the indoor exhibition scope visitors can visualize, manipulate and browse exhibition information. Some usage examples of AR in museums and galleries are audio augmented guides [2], head mounted displays-based applications [3][4], interactive museum guides over mobile devices [5][6][7], screen visualization-based systems [8] and video projectors-based systems [9].

The goal of the AR system presented in this paper is to provide an improved experience to the visitors by offering a better interaction with the exhibition contents. In order to do so, the system must immerse visitors into the augmented environment using an interface which allows a natural interaction with these contents. Moreover, the system has to provide to the visitors the educational information of the exhibition in a simple and enjoyable way.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 576–579, 2011.

2 System Description

The AR system presented in this paper is composed of a Region of Valencia map, 7 pointers and several hosts which can work in an independent way or synchronized on the same exhibition environment. Each host is formed of a PC, two visualization screens and one Kinect camera [10]. The map dimensions are 3x5 meters and visitors can walk on it. This map has a marker printed on it which is used to compute the spatial reference of the system. This map also contains 70 different hot points grouped by colors. Each color refers to a different subject of the Region of Valencia heritage and there is one pointer associated to each subject. Each pointer has a marker printed on it which is used to calculate its position on the map.

The system can be split into 4 subsystems. The communications subsystem communicates all hosts using a Client-Server model. It performs the synchronization between them over the same exhibition environment. The logic subsystem manages the system operation. It selects the subject associated to the active pointer in the system and identifies the hot point when the pointer is placed on it. The tracking subsystem obtains the position and orientation of the elements of the system: map, pointers and Kinect camera. The visualization subsystem shows augmented information over the real exhibition environment content. Figure 1 shows the different subsystems defined.

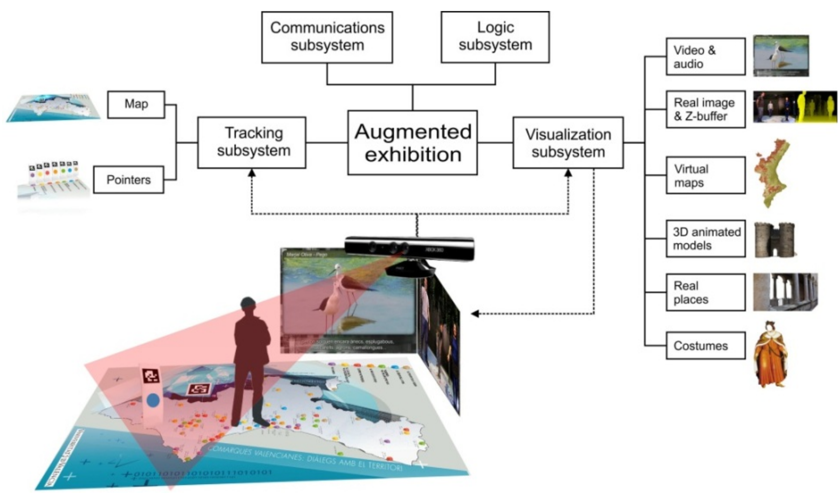


Fig. 1. Augmented reality system components

In order to mix real exhibition content and augmented reality information, initially the Kinect camera captured information is painted. This image is a reflection of the real exhibition environment. Using the depth map information given by the Kinect camera, the values of the z-buffer [11] for each image pixel are calculated. In this way, the elements of the image, visitors included, can be placed correctly in the virtual scene. Finally, the virtual elements are painted taking into account the z-buffer information, performing a correct occlusion between real (visitors and exhibition room) and virtual (3D models, maps, etc.) worlds. This process can be observed in Figure 2.

Fig. 2. Augmented reality scene composition using depth information

3 Interaction Process, Augmented Map

The visitors can walk on the large real map, as if they were traveling through the region of Valencia, using the different pointers to select the place they want to visit. When a visitor catches a pointer the system shows a multimedia resource of the subject selected on one screen. At the same time, on the other screen, an augmented visualization of a subject map replaces the real one. Thereafter, visitors can select a hot point of the selected subject on the map just "travelling" to the desired location and placing the pointer on it. Once the point is selected, a multimedia resource associated to it is showed on one screen, while on the other screen, a virtual 3D object is showed in the augmented visualization. In the same way, visitors can choose another hot point or change the subject selected by leaving the actual pointer and caching a new one. There are 3 different kinds of virtual 3D objects that can be displayed in the augmented visualization: 3D animated models which represent buildings or objects related to the hot point, virtual scenes created using images of real places and costumes to "dress up" the visitor.

4 Conclusions

The AR system presented in this paper has been used in the exhibition "Valencia, tierra de comarcas: Diálogos con el patrimonio". The system has worked properly and has been tested by over 1000 visitors. The visitor's opinions about the system were positive; they learned to use the system very fast, achieving a high level of satisfaction and entertainment.

According to the visitor's experiences, the goals have been met. The AR system was able to provide an improved experience to the visitors, supporting the learning process with the multimedia resources and virtual 3D elements associated to the hot points of the exhibition. Furthermore, the interface based on selection pointers and hot points located on the map was used by the visitors naturally.

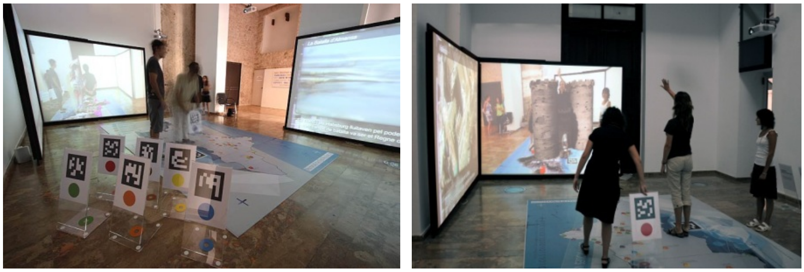


Fig. 3. 7 Pointers and exhibition set (left); visitors using the system (right)

Acknowledgements. The AR system development has been financed by the general foundation of the University of Valencia, with the collaboration of the "Vicerrectorado de Arte, Cultura y Patrimonio" and the Linguistic and Geography department.

References

1. Azuma, R.: A Survey of Augmented Reality. Presence: Teleoperators and Virtual Environments 6(4), 355–385 (1997)
2. Bederson, B.B.: Audio augmented reality: a prototype automated tour guide. In: Mack, R., Miller, J., Katz, I., Marks, L. (eds.) ACM Conference on CHI 1995. ACM Press, New York (2003)
3. Mase, K., Kadobayashi, R., Nakatsu, R.: Meta-Museum: A supportive Augmented-Reality Environment for Knowledge Sharing. In: Conference VSMM 1996, pp. 107–110. IEEE Computer Society Press, Los Alamitos (1996)
4. Abate, A.F., Acampora, G., Ricciardi, S.: An Interactive Virtual Guide for the AR Based Visit of Archaeological Sites. Journal of Visual Languages & Computing (2011)
5. Bruns, E., Brombach, B., Zeidler, T., Bimber, O.: Enabling mobile phones to support large-scale museum guidance. IEEE Multimedia 14(2), 16–25 (2007)
6. Miyashita, T., Meier, P., Tachikawa, T., Orlic, S., Eble, T., Scholz, V., Gapel, A., Gerl, O., Arnaudov, S., Lieberknecht, S.: An augmented reality museum guide. In: 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, pp. 103–106 (2007)
7. Explorer: American Museum of Natural History, `http://www.amnh.org/apps/explorer.php`
8. Bimber, O., Raskar, R., Inami, M.: Spatial Augmented Reality. In: SIGGRAPH 2007 Course 17 Notes (2007)
9. Mistry, P., Maes, P., Chang, L.: WUW – Wear Ur World – A Wearable Gestural Interface Spatial Augmented Reality. In: ACM CHI 2009, Boston (2009)
10. Kinect Home Page, `http://www.xbox.com/es-ES/kinect/`
11. Breen, D.E., Whitaker, R.T., Rose, E., Tuceryan, M.: Interactive Occlusion and Automatic Object Placement for Augmented Reality. Computer Graphics Forum 15, 11–22 (1996)

Natural Interaction without Marks

Carina S. González-González*, David Cabrera-Primo,
Melvin Gutierrez, and Jose Sigut-Saavedra

Department of System's Engineering and Automatic and Computer Architecture,
University of La Laguna, Tenerife, Spain
cjgonza@ull.es

Abstract. In this paper we present a natural interaction system that simulates an interactive mirror behavior where a subject or object can appreciate in real time the effects of external agents over themselves and the causes or actions that trigger these effects. It is a low cost system and easy to use, personalize and configure, which makes it extensible to different operating sectors, especially on the education area for interactive demonstrations. The system does not use marks and realize the detection and projection of effects in real time. For the system development a technology was invented and developed that originated the patent request ES200901210.

Keywords: HCI, Natural Interaction, Adaptive interfaces, Augmentative reality.

1 Introduction and Proposal

Currently there are various devices and processes for content generation, animation or visual effects within an area bounded by the silhouette of an object or subject. These devices and processes allow the user to know, for example, the appearance of the object or the same user when certain conditions are changed. Thus, an individual can see how you feel, for example, a certain glasses without the need for physically the same, but shown on a screen would look if you use them. Similarly, are used for interaction with different types of hair, or clothing. Using this type of system has many uses in automotive to observe the internal or external appearance of a vehicle color, trim, having only a single physical element. However, these devices have the disadvantage that visual effects displayed or do not fit subjects or objects when they are in motion or for motion detection is necessary that the subject or object has certain clothing marks as a particular color, body sensors, etc. [1,2,3, 4,5,6]. Other systems use a database from images that provide reference points associated with the subject, using external devices such as "marks" the subject that identify the area where the act [7,8,9]. These systems are used for visualization of interventions with augmented reality [10]. In this case, as well as stereoscopic vision requires a preprocessing of the images again makes reference to the object / subject for gesture recognition.

The above works differs from our proposal both in the approach to transform a skeleton outline, identifying body parts and creating a vector that analyzes the amount

* Corresponding author.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 580–583, 2011.

of movement considering these parts, as in the use of techniques and filters to obtain the proposed solution (linear Hough transform, Kalman filter to track time Fourier transformer for extracting and frequency characteristics of movement, etc.).

Our proposal is based in a system that allows the application of visual effects in objects or subjects static or in movement in real time without need of marks, archiving a natural interaction between the object or subject and the input and output devices for which computer vision techniques are used to detect, analyze and apply the different visual effects on the images in real time.

As a technical advantage should be mentioned that the interaction between objects or subjects and the input and output devices its done without marks on the objects or subject, so it is non-invasive and non-intrusive. Thus the object or subject interacts with its own body and with its natural movements if it is moving.

This technology can be applied to generate content where an external observer, or the object/subject itself appreciate how act external agents or actions realized in/with/on the body. Example given it is possible to generate visual effects related with the clothes, so a subject see the aspect of a clothing on the figure and how it adjust to the movement the subject realize. It can also show aspects related to hair, inner anatomy of the human body or any other event, graphic effect or animation desired. Said events graphic effects or animations are applied on the body in real time.

For the system to function the application of different algorithms that are part of each filter implemented in the real time image processing and movement detection module is necessary. There are two fundamental filter types: detection filters and projection filters. Where the "detection" filters are: a) a filter that allows the detection of the zone occupied by the subject and b) a filter of movement that allows the detection of the subject movements/actions. On the other hand, the projection filters that had been developed are: a) halo effect, b) desaturated effect, c) show animation in subject or object zone and movement tracking effect.

All filters are initialized with a series of parameters that can vary according to conditions of light, size and effect to achieve. It is possible to include from 1 to n filters and configure them. For visualization, they are again scaled to the adequate resolution for the used output device in each case.

2.1 Case of Application: ACTIMEL Functionality Demonstration

The system allows the application of visual effects on objects or static or moving subjects in real time without the need for marks.

The main objective in the case of ACTIMEL product demonstration is to show how the digestive system of a person before and after drinks a given liquid substance, so the person can see in the output device an animated visual effect of the body insides. Then the subject is placed in the scenario and positioned in front of the camera taking as reference the display in front of the LCD projector that the subject interprets as if it were a mirror. The subject image is projected. Then apply the detection subject filter, by cutting the background figure. This mask is used as a reference subject to filter visual effect "desaturated", creating the impression in the subject of a loss of color, which is imaginary interpreted as a possible dehydration. Then the subject must perform the action of drink the liquid, for which grabs a bottle and brings it to the mouth. This action is detected by the movement mask filter. Drink action is interpreted by the start and end pre-selected positions points in the frames, as noted above in the explanation of the filter operation. When detecting the action, it triggers the application of the visual effect animation playback filter within the subject area and motion tracking. Within the area of reproduction, the desired animation is shown, for example, a ball falling through the body and its expansion into of the body (drop down the throat to the intestine and there are expanded), as shown in Figure 1. The subject can move on stage and the animation will follow your movement and will readjust to its size thanks to the previous filter. When this animation ends, it triggers the next filter, the "halo" visual effect, where a colored halo draws the silhouette of the subject with predefined characteristics, such as shown in Figure 2. At all times the subject can move in the scenario and the effects are adjusted to their movements. This is because the image analysis and processing are made frame by frame in real time.

Fig. 1. Projection of animation effect with automatic adaptation to the user's characteristics.

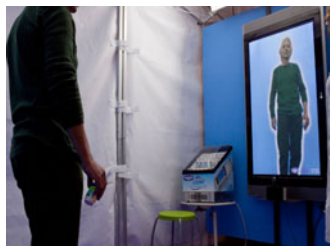


Fig. 2. Effect "halo" projected on the user

Validity was determined from the laboratory to the implementation of the system on the market with different user profiles (child, adult, senior) and different effects on their image. The sample consisted of 34 individuals aged between 10 and 60 yr, 12 women and 22 men. The tests gave highly satisfactory results in terms of robustness of the system and user satisfaction regarding its use. Moreover, the Mirror Effect system was put on the market in May 2009 at the Mall ALCAMPO La Villa, La Orotava, and Tenerife. It is estimated that an average of 100-130 people made use of the system per day during the months in operation. Between May 2009 and July 2010 the system was taken to various shopping malls in the island of Tenerife (Carrefur, ALCAMPO La Laguna, etc..) and on the island of Gran Canaria, obtaining similar results in their reception. Moreover, the company found DANONE Canary increases Actimel product sales during the promotional periods.

3 Conclusions

In this article we have presented a system that allows the application of visual effects on objects or static or moving subjects in real time without marks through the following detection filters: a) a filter object that allows the detection of the area occupied by the object and b) a motion filter that allows detection of movements or actions developed by the object.

This system has been funded by a company business' contract and the Canary DANONE research team from the Department of Systems Engineering and Automation at the University of La Laguna and was selected for successful business transfer through the Network of Spanish University Transference Offices (OTRIs) in 2009.

References

1. Zhou, F., Duh, H.B.-L., Billinghurst, M.: Trends in Augmented Reality Tracking. In: Interaction and Display: A Review of Ten Years of The IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR, 7th IEEE/ACM International Symposium on Mixed and Augmented Reality, ISMAR 2008 (2008)
2. Auer, T., Pinz, A.: Building a hybrid tracking system: Integration of optical and magnetic tracking. In: IWAR 1999, pp. 13–22 (1999)
3. Azuma, R.T.: A survey of augmented reality. Presence: Teleoperator and Virtual Environments 6(4), 355–385 (1997)
4. Azuma, R.T., Baillot, Y., Behringer, R., Feiner, S., Julier, S., MacIntyre, B.: Recent advances in augmented reality. IEEE Computer Graphics & Applications 21(6), 34–47 (2001)
5. Azuma, R.T.: A survey of augmented reality. Presence: Teleoperator and Virtual Environments 6(4), 355–385 (1997)
6. Klein, G., Murray, D.: Parallel Tracking and Mapping for Small AR Workspaces. In: The IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR 2007, pp. 225–234 (2007)
7. Klinker, G., Reicher, R., Brugge, B.: Distributed user tracking concepts for augmented reality applications. In: ISAR 2000, pp. 37–44 (2000)
8. Skrypnyk, I., Lowe, D.G.: Scene modeling, recognition and tracking with invariant image features. In: The IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR 2004, pp. 110–119 (2004)
9. White, S., Lister, L., Feiner, S.: Visual hints for tangible gestures in augmented reality. In: The IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR 2007, pp. 47–50 (2007)
10. Riva, G., Gamberin, L.: Virtual reality as telemedicine tool: technology, ergonomics and actual applications. Journal Technology and Health Care Archive 8(2) (July 2000)

NAVI – A Proof-of-Concept of a Mobile Navigational Aid for Visually Impaired Based on the Microsoft Kinect

Michael Zöllner, Stephan Huber, Hans-Christian Jetter, and Harald Reiterer

University of Konstanz, Germany
{michael.zoellner,stephan.huber,hans-christian.jetter,
harald.reiterer}@uni-konstanz.de

Abstract. We present a proof-of-concept of a mobile navigational aid that uses the Microsoft Kinect and optical marker tracking to help visually impaired people find their way inside buildings. The system is the result of a student project and is entirely based on low-cost hard- and software. It provides continuous vibrotactile feedback on the person's waist, to give an impression of the environment and to warn about obstacles. Furthermore, optical markers can be used to tag points of interest within the building to enable synthesized voice instructions for point-to-point navigation.

Keywords: Vibrotactile User Interface, Navigation User Interface.

1 Introduction

Wayfinding is a cognitive demanding, error-prone task for visually impaired persons. It is described by Long and Hill as "the process of navigating through an environment and travelling to places by relatively direct paths" [8]. As such, it depends heavily on both "sensing of the immediate environment for obstacles and hazards" [9] and "navigating to remote destinations beyond the immediate perceptible environment" [9]. These two aspects of wayfinding are also called *micro-navigation* and *macro-navigation* [10]. Although both aspects are part of successful wayfinding, there are only few systems that try to integrate them together (e.g. [1, 7]). These systems typically try to substitute their users' limited or missing visual capabilities by giving helpful feedback via different sensory channels. Sound and speech are probably the most frequently used sensory channels in this domain. However, as environmental auditory signals are very important for persons with limited sight, acoustic feedback must be used carefully and is not appropriate in all cases. Especially when it is used for continuous feedback - which is very important in micro-navigation scenarios such as obstacle detection - acoustic feedback can become annoying and distract users from the natural aural environment. In macro-navigation scenarios, speech output - similar to that of car navigation systems – seems to be a reasonable choice. However, as different users interpret these signals differently, it is not easy to give unambiguous navigation instructions. Even worse, workload may differ if navigation instructions are given by sighted persons instead of visually impaired persons [2]. It "seems to be a hopeless situation for the ability of language to convey spatial information" [6]. An alternative is vibrotactile feedback that is used for navigation purposes in various ways, ranging from vibrating headbands [4] to shoulder tapping systems [12]. As

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 584–587, 2011.

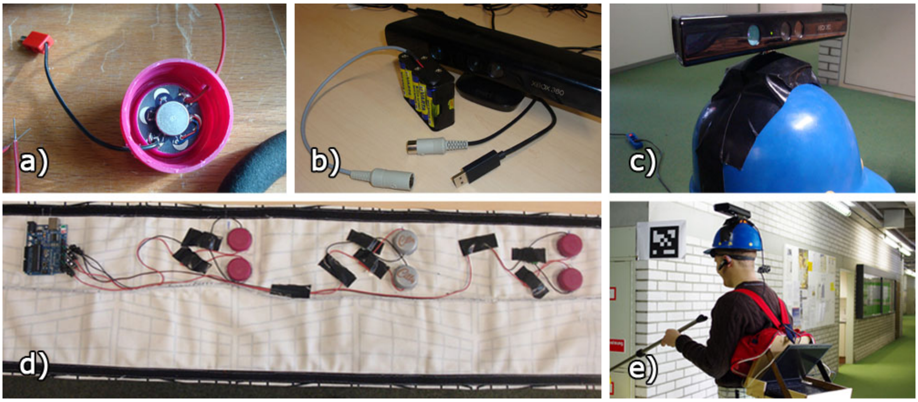


Fig. 1. (a) LilyPad vibe board in a plastic cap. (b) Mobile Kinect camera with battery pack. (c) Kinect helmet. (d) Vibrotactile Waist Belt. (e) Complete setup with backpack.

vibrotactile signals are less disruptive than sound or speech, they are an appropriate alternative to give continuous feedback. Thus, our system employs a hybrid approach that augments synthesized speech instructions for macro-navigation with vibrotactile feedback for micro-navigation using only low-cost and self-made technology.

2 System Description

The initial design of our proof-of-concept was created during HCI coursework at our university. Our design goal was to build a system that enables and/or facilitates both micro-navigation and macro-navigation. For this we utilized the Microsoft Kinect sensor, a vibrotactile waist belt built with Arduino LilyPad[1] vibe boards and a simple backpack construction that carried the laptop and enabled quick debugging (Fig.1 e).

For detecting the immediate surroundings, we reversed the standard operating principle of the Kinect. Instead of a static Kinect that tracks moving objects, we track the static environment with a moving head-mounted Kinect (Fig.1c). To power the mobile Kinect we use a 12V battery pack that lasted for about 5 hours during our tests (Fig.1 b). The vibrotactile output is provided by a waist belt that contains three pairs of Arduino LilyPad vibe boards (Fig.1a). These are fixed into plastic bottle caps to amplify the perceived vibration (Fig.1d). The speech output is provided by an ordinary Bluetooth headset for mobile phones (Fig.1e).

Micro Navigation – In contrast to Lee et al., who use vibrotactile signals to encode navigation instructions [7], we use them to inform users of persons or other obstacles in their way, thereby enabling secure micro-navigation. Erp et al. showed that encoding distances on a vibrotactile display is realized best with temporal patterns instead of changes in frequency or amplitude [5]. Based on several informal evaluations with different distance signals, we decided for a simple yet robust distance encoding with just three different outputs: no output meaning no obstacle,

[1] `http://hlt.media.mit.edu/?p=34`

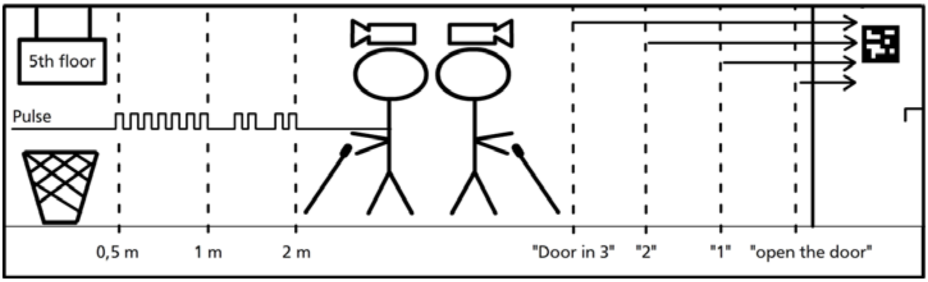


Fig. 2. Vibrotactile cues and navigation instructions depend on the distance to the obstacle (left side) or marker (right side)

a 500ms pulsed vibration for signaling obstacles between 1-2m distance and a continuous vibration for obstacles closer than 1m (Fig. 2). Erp et al. also investigated that spatial information is encoded most effectively by applying the vibrotactile signals to different body locations [5]. Thus we located the vibe boards on the left, right and center of the person's waist to indicate the direction in which an obstacle was detected. Unlike Cardin et al. [3], who detect only the closest obstacle, we are able to detect the closest obstacles in the left, right and center region of the Kinect's "field of view". The signal processing software and the control of the vibe boards is implemented with C#/.NET and uses a managed wrapper of the OpenNI framework[2]. Obstacles in each of the image's regions are identified via a depth histogram. We assume that nearby and large objects are potentially harmful to the person. Thus, the obstacle detection algorithm moves a depth window of 120mm size from left to right (respectively near to far) over the histogram and stops, if the pixel area of that depth window exceeds a certain threshold area (approx. 4% of a region). The average depth value of the current depth window is then mapped to the pulse of the vibe board.

The overall minimalistic design of the vibrotactile output (3 output channels with 3 different signals each) is based on several iterations with informal experiments, where users felt more comfortable and performed better with less sensory information. We therefore chose to minimize the output's complexity thereby also minimizing the users' cognitive load.

Macro-Navigation – We use synthesized voice as feedback mechanism for macro-navigation. Since navigation instructions are very short (e.g. "turn left") and don't occur continuously, they only slightly interfere with environmental sounds. Our low-cost solution uses printed augmented reality markers (Fig.1e) that can be used to tag points of interest on the desired route. We detect the markers via Kinect's RGB camera using a managed wrapper of the ARToolkitPlus[3]. By integrating depth information of the Kinect, we facilitate different navigation instructions based on the person's distance to the marker. For example, when walking towards a door, the instructions will be as follows: "Door in 3", "2", "1", "Open the door" (Fig. 2). Our approach has the drawback that every route has to be signposted individually. While the effort to create and maintain such a relative positioning system is higher compared to a map-based absolute positioning system like GPS, it has the essential advantage of being available inside buildings.

---

[2] `https://github.com/kobush/ManagedOpenNI` (Retrieved June 2, 2011).
[3] `http://code.google.com/p/comp134artd` (Retrieved June 2, 2011).

3 Future Work

In this work we presented a system that facilitates micro- and macro-navigation for visually impaired persons. We plan to iteratively improve this proof-of-concept in the future: (1) We observed that users don't walk straightforward and frequently don't approach markers from the front but from varying angles. This rendered some navigation instructions useless, as they were designed for frontal approaches. Future work should adapt navigation instructions to the angle from which the user is approaching. (2) Like Ram and Sharf [11] we would like to discriminate between persons and inanimate objects. This would help to reduce the amount of system output while moving in a crowded environment and could prevent users from sensory and cognitive overload. (3) Finally, we plan to conduct a formative user study with visually impaired users to evaluate navigation instructions and vibrotactile signals. A further summative study could then reveal how NAVI compares to the conventional white cane in terms of navigation effectiveness, efficiency and cognitive load.

References

1. e-adept: Electronic assistance for disabled and elderly pedestrians and travelers (2011), retrieved June 2 from `http://www.eadept.se`
2. Bradley, N.A., Dunlop, M.D.: An experimental investigation into wayfinding directions for visually impaired people. Personal Ubiquitous Computing 9, 395–403 (2005)
3. Cardin, S., Thalmann, D., Vexo, F.: A wearable system for mobility improvement of visually impaired people. Vis. Comput. 23, 109–118 (2007)
4. Cassinelli, A., Reynolds, C., Ishikawa, M.: Augmenting spatial awareness with haptic radar. In: Cruz, I., Decker, S., Allemang, D., Preist, C., Schwabe, D., Mika, P., Uschold, M., Aroyo, L.M. (eds.) ISWC 2006. LNCS, vol. 4273, pp. 61–64. Springer, Heidelberg (2006)
5. Erp, J., Veen, H., Jansen, C., Dobbins, T.: Waypoint navigation with a vibrotactile waist belt. ACM Transactions on Applied Perception 2(2), 106–117 (2005)
6. Franklin, N.: Language as a means of constructing and conveying cognitive maps. In: Portugali, J. (ed.) The Construction of Cognitive Maps. GeoJournal Library, vol. 32, pp. 275–295. Springer, Netherlands (1996)
7. Lee, J., Choi, E., Lim, S., Shin, B.: Wearable computer system reflecting spatial context. In: Proc. IWSCA 2008, pp. 153–159. IEEE Computer Society, Washington, DC (2008)
8. Long, R.G., Hill, E.W.: Establishing and maintaining orientation for mobility. In: Wiener, W.R., Blasch, B.B., Welsh, R.L. (eds.) Foundations of Orientation and Mobility, 2nd edn., ch. 2, pp. 39–59. AFB Press (1997)
9. Loomis, J.M., Golledge, R.G., Klatzky, R.L.: Gps-based navigation systems for the visually impaired. In: Barfield, W., Caudell, T. (eds.) Fundamentals of Wearable Computers and Augumented Reality, L. Erlbaum Associates Inc., Hillsdale (2000)
10. Petrie, H.: User requirements for a gps-based travel aid for blind people. In: Proc. of the Conference on Orientation and Navigation Systems for Blind Persons, RNIB (1995)
11. Ram, S., Sharf, J.: The people sensor: A mobility aid for the visually impaired. In: Proc. ISWC 1998, p. 166. IEEE Computer Society, Washington, DC (1998)
12. Ross, D.A., Blasch, B.B.: Wearable interfaces for orientation and wayfinding. In: Proc. ASSETS 2000, pp. 193–200. ACM, New York (2000)

OntoCompo: A Tool to Enhance Application Composition

Christian Brel, Anne-Marie Dery-Pinna,
Philippe Renevier-Gonin, and Michel Riveill

I3S Laboratory (Université Nice-Sophia Antipolis / CNRS),
930 route des Colles, BP 145,
06903 Sophia Antipolis Cedex, France
{brel,pinna,renevier,riveill}@polytech.unice.fr

Abstract. Mash-ups emerged through the web 2.0 to juxtapose several applications and use them together. The next step after juxtaposition is the composition of existing applications to build a new one. A solution of this being born need is the reuse of parts from formers applications. To perform this composition and reuse in an easy and comfortable way, we propose a tool based on several extensions of selection to help the developer during his composition.

Keywords: Application Composition, Semantic Annotation, CBSD, UI, task model.

1 Introduction

The advent of web 2.0 and the apparition of a lot of "applications stores" introduce implicitly new needs for users and developers faced to this set of applications disposed on the web. Mash-up solutions for example allow them to juxtapose several applications and use them together. They can have ideas for new functionalities creating a new application combining existing ones. Adapting applications to users' requirements may be done through composition of applications. Tools for composing former applications (and probably corresponding source codes) should introduce developers' comfort and a reduction of the time-to-market for new applications by recycling former applications.

In this paper, we present the tool OntoCompo dedicated to easily realize new applications by composition of their User Interface. This tool deals with component-oriented applications respecting a separation in two parts: the User Interface (UI), visible and well-known part of the application and the functional core (FC), underground part of the application. Due to this clear separation, the composition process lets the possibility to the developer to build the new application selecting, extracting and positioning UI part of former applications, one after another [**2**]. So we focus on the connections between UI, FC and tasks. We consider that a composition driven by a checked selection is a guarantee to preserve the global consistency of the final application. So we choose to help the developer for broadening selection.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 588–591, 2011.

In the next section, we describe related works and we underline our originality. Then we present the hypothesis of our work and our tool for application composition.

2 Related Works

As we aim at composing applications by manipulating their UI, we have to decompose UI, i.e. describe UI in order to deal with sub-parts of former UI. The description of an UI both involves (1) description of its structure (like UIML [1], ALIAS [7], UsiXML [5] or MARIA [8]) and (2) the spatial positioning of these components (like in different layouts used in the UI toolkits).

To manipulate applications in order to compose them, there are currently three main approaches: (i) the composition could be triggered by the functional part as in [7], (ii) the composition could be triggered by the users' goals (i.e. tasks) as in [8] and (iii) the composition could be triggered by the UI as in [4].

Each trigger addresses a specific problem of composition: presentation and layout considerations at the UI level, behavior of the application at the functional level, users' needs at the task level. These works do not reuse complete architecture of the former applications. Either they compose and reuse UI as first concern without any consideration of the links between UI and the functional part either their first concerns are functionality or task and provide the new application by (re-)generating UI.

Our originalities are (i) to consider links between UI, tasks and functionalities, (ii) to lead the developer by suggesting him and asking him about elements to keep for aiming at composition consistency and (iii) to reuse existing UI in order to preserve former developments, former designs and former practices. Our tool, OntoCompo, helps the developer of application for reusing existing applications to constitute his new one. We purpose the developer to select UI elements he wants to keep and suggest him extensions for his selection in order to obtain a new functional application after composition.

3 Hypothesis on Former Applications

To be able to reuse elements of the former application, we need a software organization authorizing selection, extraction and rejigging of such elements. We opt for applications developed with FRACTAL components [3]. For reusing of former applications parts, we use: (i) component-based software development to manipulate functionality assemblies and (ii) component-based UI with Java Swing JComponent encapsulated in FRACTAL component in order to manipulate concrete UI parts. Applications are not expected to be provided with sources. Indeed FRACTAL components are seen like black box and inputs and outputs software *interfaces* are available. To reuse existing applications, our hypothesis is to let the developer doing composition through the interfaces of applications. So, our approach is to enhance links and to extend connections between UI elements and Functional Core elements. That strengthening is based on the Task Model (TM). We use semantic annotations (using OWL Light[1] language) for the description of applications. So the OWL Light

[1] OWL Web Ontology Language. `http://www.w3.org/TR/owl-features/`

description includes the description of the task model (an OWL representation of CTT [6]), the description of the UI elements (an OWL representation of MARIA [8]) and their layout and the description of functionalities. The OWL Light description also includes links between tasks and functionalities, links between tasks and UI elements, links between functionalities and the *concrete* FRACTAL component, links between UI elements and the *concrete* FRACTAL component.

4 Composing Thanks to Extensions of Selection

The simple selection of a part of an application is the direct manipulation. By a click on an UI element, the developer can select it in order to extract it later. Selected UI elements are graphically highlighted. That simple selection is extended for performing complex selections or aiming at verifying consistency.

First, there is ***the layout extension***. With the height toggle buttons for selected extension directions, the developer has the possibility to broaden the selection. SPARQL[2] queries are parameterized with the current selection and with each chosen directions. Such a query returns the relevant fractal component identifiers.

Secondly, there is ***the (container) parent extension***. It's also about queries layout of application to obtain the parent container of last selected UI component in current selection. This extension allows the developer to be more efficient on his selection of all elements in a container potentially "hidden" by its contents.

Thirdly, there is the ***task extension***. Each UI element is linked with a task described with semantic annotations. From the last selected component, we use SPARQL queries to obtain the task linked to it. From each returned tasks, we query semantic annotations to obtain all UI elements linked with this task.

Finally, there is ***the functionality extension***. UI elements are directly linked to functionality but also through tasks. Since a task may be connected to several functionalities, it is possible to extend the selection to each part of the application by following these links. We start with selected UI elements. Thanks to SPARQL queries, we go back "up" to related tasks and then "up" to related functionalities. From these functionalities, we go back "down" to UI elements.

Each of these extensions can be activated by the developer. He is free of combination between all proposed extensions. To help him and to lead him towards to a coherent composition, we develop a help selection. This help is a guide for the developer during all selection process. For each UI element, several questions suggest to the developer different possibilities for extending his selection. The developer can partially or fully use that help (guided by tasks and/or by functionalities and/or by layouts) to perform his selection.

5 Conclusion

To conclude, with OntoCompo, we provide a solution to compose application from a manipulation of UI. We help the developer during the composition with all proposed selection extensions. Those selection extensions based on suggestions to enhance the

[2] SPARQL Query Language for RDF. `http://www.w3.org/TR/rdf-sparql-query/`

reused part of former applications lead to an usable application. The developer being able to choose his entry point (UI layout, functionalities or tasks) to perform his extensions, we are now planning developer evaluation to validate the different extension. Once that evaluation performed, we will work on a new step in the composition process about merging application elements (UI elements or functionalities).

References

1. Abrams, M., Phanouriou, C., Batongbacal, A., Williams, S., Shuster, J.: UIML: An appliance-independent XML user interface language. In: Proceedings of the 8th World Wide Web Conference (WWW), pp. 617–630. Elsevier, Amsterdam (1999)
2. Brel, C., Renevier-Gonin, P., Occello, A., Pinna-Déry, A.-M., Faron-Zucker, C., Riveill, M.: Application Composition Driven By UI Composition. In: Proceedings of the Human Computer Software Engineering 2010 (HCSE 2010), IFIP International Federation for Information Processing, 2010. LNCS, pp. 198–205 (October 2010)
3. Bruneton, E., Coupaye, T., Leclercq, M., Quéma, V., Stefani, J.-B.: The FRACTAL component model and its support in Java. Software: Practice and Experience 36, 1257–1284 (2006), doi:10.1002/spe.767
4. Fujima, J., Lunzer, A., Hornbæk, K., Tanaka, Y.: Clip, Connect, Clone: Combining Application Elements to Build Custom Interfaces for Information Access. In: Proceedings of UIST 2004, Santa Fe, NM, pp. 175–184 (2004)
5. Limbourg, Q., Vanderdonckt, J., Michotte, B., Bouillon, L., Florins, M., Trevisan, D.: Usixml: A user interface description language for context-sensitive user interfaces. In: AVI 2004 Workshop Developing User Interfaces with XML: Advances on User Interface Description Languages, UIXML 2004, pp. 55–62 (2004)
6. Mori, G., Paternò, F., Santoro, C.: Ctte: Support for developing and analyzing task models for interactive system design. IEEE Transactions on Software Engineering, 797–813 (2002)
7. Occello, A., Joffroy, C., Pinna-Déry, A.-M., Renevier-Gonin, P., Riveill, M.: Experiments in Model Driven Composition of User Interfaces. In: Eliassen, F., Kapitza, R. (eds.) DAIS 2010. LNCS, vol. 6115, pp. 98–111. Springer, Heidelberg (2010)
8. Paternò, F., Santoro, C., Spano, L.D.: Maria: A universal, declarative, multiple abstraction level language for service-oriented applications in ubiquitous environments. Computer-Human Interaction (TOCHI) 16 (2009)

Personal Chart: Health Data Logging Made Easy with a Mobile Device

Mikko Paldanius[1], Anu Lehtiö[2],
Minna Karukka[1], and Pertti Huuskonen[2]

[1] Nokia Research Center, Yrttipellontie 10, 90230 Oulu, Finland
[2] Nokia Research Center, Visiokatu 1, 33720 Tampere, Finland
firstname.surname@nokia.com

Abstract. Many devices are still unconnected. We apply mobile imaging to log measurements from personal health devices. Such devices sometimes offer wired and wireless links, but they suffer from many problems (setup, breaking connections, non-mobility). We propose to use instead ubiquitous mobile phone cameras to capture the measurements and store them for further viewing and follow-up. In this paper we discuss the principle, the prototype, the user study and initial conclusions of this approach.

Keywords: Healthcare, human factors, optical character recognition.

1 Introduction

Today's personal health data measurement devices (e.g. blood pressure meters, glucose monitors) are still to a large degree disconnected. Their users need to keep track of the measurements manually, and transfer the measurements to logbooks or analysis software. This obviously is tiresome and creates a bottleneck for automated health monitoring. The same issues have existed a long time in clinical environments. With decreasing staff and increasing amounts of data, medical workers would appreciate automated data logging from their instruments.

Many clinical and some consumer devices do feature wired or wireless links for transferring the data to PCs and eventually to internet services. There are also devices that connect to mobile phones for the same functionality [1, 2]. However, in too many cases the users need to take elaborate steps to initiate the data transfer. Wireless technology is still difficult for people who are averse to technology: convoluted setup procedures and intermittent connections confuse them. The problem is worse for mobile use situations, where available interaction times may be reduced to seconds [3]. In earlier research e.g. gestures have been suggested for more intuitive way of pairing devices and interacting with other devices [4, 5].

We propose a different approach. Instead of relying to wireless links or cables, we use mobile imaging for data transfer. A large fraction of mobile phones in the market already include suitable hardware for the purpose. The phone camera can take a photo of the display of a personal measurement device, interpret the measurement reading through OCR (Optical Character Recognition), and store the data for later retrieval or uploading into a web service. This gives people a fast, convenient way to keep track

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 592–595, 2011.

of their health data. It avoids problems related to wireless links, empowering people to use an interaction method they already know: point and shoot. In this paper we introduce and evaluate an implementation of this approach on mobile phones.

2 Personal Chart Application

Personal Chart for Nokia N900 is an application for detecting blood pressure values from a measuring device. Personal Chart has two main functionalities: capturing blood pressure values and viewing measurement history. In addition, users can share their personal measurement information via Bluetooth or email. After the application is launched for the first time, a user is asked to create a profile. Profile is made up of name, gender and weight. The application enables using several profiles.

Capturing is launched by clicking the camera icon on the start view. A user is guided to point camera to measuring device and match the live view screen marks with the display of measuring device. Values can be entered manually by pressing skip tracking icon. Once the values are detected from the live view, the confirmation view is opened. The user can manually alter the values in case of misdetection. In the confirming view the user can save values or backtrack to capturing mode.

Measurement history can be viewed as a table layout and a graphic line chart. In the table view, a user can see all the values in chronological order. In the chart view, presented in figure 1, measurement history is portrayed in a line graph according to measured value and time. The user can choose which lines are drawn in the graph by toggling the check boxes. Saved data can be shared via email or Bluetooth. Shared file is a CSV file (Comma Separated Values) that opens in almost all spreadsheets, including Microsoft Excel and OpenOffice Calc.

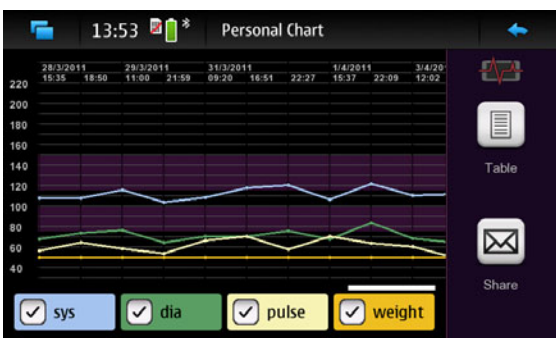


Fig. 1. Chart view of the Personal Chart application

3 Evaluation

The main goal of our study was to evaluate the general acceptability of the application concept and the applied technology. Basic requirements for participants were chosen to be the age of 25-55 years, experience in smart phones and the use of touch screen devices. It was preferred that both sexes are represented. Evaluation was based on user testing session and voluntary online discussion. Originally the interviews and the discussions were in Finnish.

3.1 User Testing

The total amount of 50 users was considered to be sufficient to give reliable and valid results. Each participant took part in 30 minute user testing session, 31 of them were male and 19 female. During the test the users were given 3 tasks, after performing each task they answered open-ended questions and graded statements related to that. After the session, users answered open-ended questions concerning the application as a whole and filled in the UX -questionnaire.

The first flow of tasks were to create a user profile, check how the used device is selected, to measure blood pressure, to transfer the readings to mobile device first by tracking then manually and save them. Creating a profile was considered easy and the add user button was clear and well placed. In addition to name, gender and weight, suggested adds for profile were age and body mass index. After creating a profile, users moved on to device selection. Choosing the device model and manufacturer was considered easy, but the name of the menu was causing confusion. OCR was the preferred method to manual inputting: Easiness, quickness, modernity and avoiding the extra step of typing were typical reasons beyond that. Manual inputting could be improved by making it more intuitive. Those few who preferred manual input had the most trouble with the lightning conditions, optical character recognition is remarkably vulnerable due to the application`s sensitivity to shadows and reflections. Saving the data was easy and not a single user had trouble with that.

The second task was to view the blood pressure data in table and chart form and to administer the graphs shown in the chart view. The chart view was preferred to the table view because of giving better overall view of development. In addition, it was argued to be visually more appealing, more illustrative and the graphs were easily distinguishable and fun. The shifting between the different views was easy and fast. In the chart view, many users were asking for some feedback from their blood pressure readings. Despite of the preference towards the chart, both views were needed. They serve a different purpose: the chart is good for getting an overall view on trends and the table gives you exact values and detailed information.

The third task was to share own blood pressure data. Users preferred Email to Bluetooth. It was considered easy and more familiar. Most likely recipients were doctor, nurse, coach and personal trainer. In addition, users wanted to send the data to themselves with the intention to storage and view the data with another device. Especially these users were interested in the possibility to use Bluetooth. The possibility to send the data forward was seen useful and add value to the use of the application. In the case of annual health examination, the data could be sent to the doctor beforehand to make analyzing and medication planning more effective. In addition, individuals could decrease the amount of visiting the doctor by measuring blood pressure unassisted and sending the data by themselves, also from abroad.

3.2 Online Discussion

After the user testing, users had an option to air one`s opinion in the discussion forum for three weeks. 11 male and 5 female took part in discussion with 5 different topics given by moderator. The best thing was argued to be the simplicity and the ability to share information. Participants defined quite unanimously that the most suitable users for this application would be the people who have to follow their blood pressure values and nursing staff who can send the values forward. Improvement and developing ideas consisted of alternative methods for data transfer, a PC client for the application and some changes for the user interface.

4 Conclusions

The Personal Chart application was considered modern and easy to use for health data logging. However, we believe that UI enhancements and more flexible character recognition would have a clear positive effect on usability of this application. In the health and well-being research, Personal Chart could create a valuable channel for collecting large amount of collaborative health data. Before that, user evaluations with doctors and nurses could be a fertile ground for future investigations.

In general, Personal Chart was seen useful and participants found many use cases for the application related to personal health management and communication with health care. It could bring true value for both doctors and patients by making their interaction with each other more effective and easier. In addition, blood pressure and blood sugar data could be easily updated from mobile phone into popular services related to exercise, nutrition and weight-watching. Since people are nowadays encouraged to take good care of themselves, an application that supports user's knowledge of his/her own health seems to be very current. In addition, this technology could suit a whole variety of applications for non-medical devices, for example with industrial maintenance.

References

1. Polar WearLink, `http://www.polar.fi/en/products/accessories/WearLink_transmitter`
2. Sports Tracker, `http://www.sports-tracker.com/`
3. Oulasvirta, A., Tamminen, S., Roto, V., Kuorelahti, J.: Interaction in 4-second bursts: the fragmented nature of attentional resources in mobile HCI. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 919–928. ACM, New York (2005)
4. Kray, C., Nesbitt, D., Dawson, J., Rohs, M.: User-defined gestures for connecting mobile phones, public displays, and tabletops. In: Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services, pp. 239–248. ACM, New York (2010)
5. Peng, C., Shen, G., Zhang, Y., Lu, S.: Point&Connect: intention-based device pairing for mobile phone users. In: Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services, pp. 137–150. ACM, New York (2009)

Psychosocial Indicators via Hand Tremor

Ted Selker, Patricia Collins, and Will Dayton

Carnegie Mellon University Silicon Valley, Moffett Field, CA 94035
{ted.selker,patricia.collins}@sv.cmu.edu, wdayton@gmail.com

Abstract. We propose hand tremor as a new type of input that can corroborate psychosocial conditions. An Android application was able to distinguish tremor variability differences between people with and without diagnosed hand tremor. Tremor measurements also corroborated self-assessment of sleep quality. Hand tremor evaluation may be a monitorable, implicit input to systems that respond to various psychosocial states. We encourage researchers to consider how interface design changes when using implicit sensors such as tremor sensing.

Keywords: psychosocial indicators, novel user interface, hand tremor, implicit input.

1 Introduction

Tremor, an involuntary, semi-periodic movement, is the most common movement disorder [4]. Most people exhibit physiologic hand tremor, a benign, low-amplitude variation in neuromuscular state [5]. Tremor may be heightened by many factors such as strong emotion, physical exhaustion, hypoglycemia, hyperthyroidism, stimulants, alcohol intake or withdrawal, or fever [1]. Other factors such as sleep [2] and smoking [7] have been shown to affect tremor significantly.

With a growing number of smartphones sold annually [3], accelerometers provide an omnipresent opportunity to measure the state of hand tremor.

Previous efforts to measure tremor to improve cellphone user interfaces have been unsuccessful. Quek, for example was not able to measure body tremor reliably to recognize when a phone was being held [8]. Other attempts at improving the user's experience by reducing the impact of tremor have had limited success [4].

The presented research demonstrates that even if tremor status might not completely indicate the user's condition, it might be part of sensor and contextual information fusion that reveals psychosocial state. The system could then provide context-specific guidance or feedback to the user.

2 Method

Most hand tremor is in the 3-6 Hz range, although certain types can reach 12 Hz [6]. We developed an application that detects tremor up to 19 HZ the HTC G1 phone, running Android 1.6. The mobile phone application records 3-D accelerometer while the subject holds the phone with an unsupported arm for 30 seconds. The number of amplitude maximums; tremor frequency; amplitude average, minimum, and

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 596–599, 2011.

maximum; period average, minimum, maximum; and standard deviation of period are stored on a micro SD card.

A pilot study included volunteers with and without tremor disorder. For each measurement, the phone's display first instructs a user to "hold the phone comfortably in your dominant hand with the screen facing you. As long as you comfortably can, extend your arm fully while keeping it perpendicular to your body. Remain as still as possible for the duration of the test."

A formal study was conducted (age ≥ 65) at a senior center. Subjects filled out a pretest survey of experience with hand tremor, took tremor measurements, and filled out the Pittsburgh Sleep Quality Index (PSQI) [9]. Subjects were given the option to have the experimenter read the sleep survey aloud and to record their responses.

3 Results

The pilot study included one subject with diagnosed tremor used the application for a few days. Several times each day this subject took a measurement, noting no tremor (N=8 times), slight tremor (N=4 times), or clearly visible tremor (N=7 times). Analysis easily distinguished between the three reported states (figure 1 and 2).

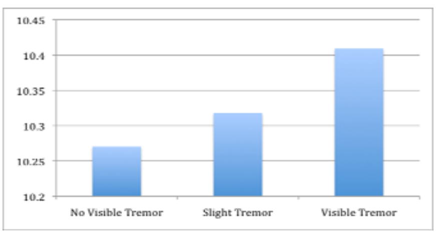


Fig. 1. Mean amplitude maximum (m/s^2)

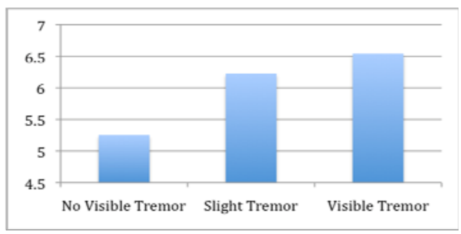


Fig. 2. Mean Tremor Frequency (Hz)

Unlike amplitude, frequency clearly distinguished the subject with known tremor. None of the non-tremor pilot subjects recorded above 5 Hz, lower than the tremor subject's mean "no visible tremor." Most subjects' hand movement was in the 3-4 Hz range. Standard deviation of period also proved to distinguish between- and within-subject comparisons (figure 3). No other pilot subject recorded a standard deviation as low as the tremor subject's.

The pretest survey and data collection protocol were administered to 19 formal study subjects interacting with a phone as a tremor sensor. Sixteen completed the PSQI sleep survey as well. Five reported a history of hand tremor disorder. As in the pilot experiment, *amplitud*e profiles did not distinguish between tremor and no-tremor subjects (Table 1). However, the two groups on frequency measurements yields a t-score of 5.06 for one degree of freedom with less than 0.1 of the result occurring by chance. Comparing the two groups on mean period standard deviation yields a t-score of 3.70, less than 0.1 that the result occurred by chance. As with pilot results, the formal study results suggest that those with hand tremor have higher frequency of tremor and lower standard deviation.

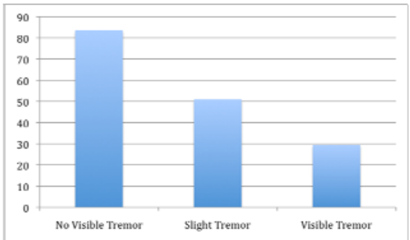


Fig. 3. Mean standard deviation tremor time period (milliseconds)

The sleep index graph (figure 4) shows a nearly linear correlation with hand tremor frequency. A correlation analysis of PSQI scores and tremor frequency yields an r-score of 0.557 with a probability of less than 0.05 that the result occurred by chance.

Table 1. Tremor amplitude and frequency between diagnosed tremor and no tremor groups

No History of Hand Tremor N= 14	History of Hand Tremor: N= 5
Mean Amplitude Maximum (m/s^2)	
10.28	**10.30**
Mean Frequency (Hz): t-score 5.06	
3.52	4.95
Mean Period Standard Deviation (ms): t-score 3.70	
138.29	77.11

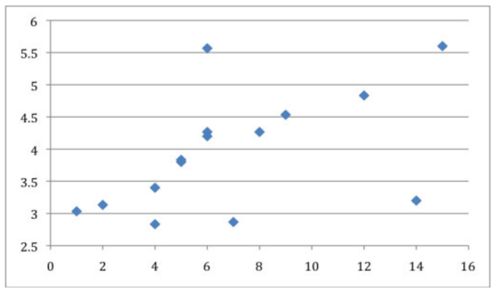


Fig. 4. Scatter plot of PSQI Scores (x-axis) and hand tremor frequency (y-axis)

4 Discussion

Smartphone measurements of tremor frequency corroborated self-report of sleep quality and tremor. The use of such an application to address the healthcare needs of those with tremor appears feasible. We report on overcoming earlier reported obstacles to using tremor to recognize tremor presence [8] and go on to distinguish diagnosed tremor and sleep quality in tremor period and standard deviation .

We examined the single factor of sleep quality and found a positive correlation between the PSQI and hand tremor. Tremor is also associated with smoking, caffeine use, physiological and emotional stress--all factors that could be examined to allow users to gain a better understanding of their health.

Our frequency domain instrument consistently distinguished tremor in subjects ranging from their 20s into their 80s. This instrument supports comparisons across people without calibration. This work only begins the story of what might be done with indirect indicators like hand tremor.

5 Conclusion and Future Work

The frequency-domain smartphone instrument could simplify neurological diagnosis and give more accurate and responsive ways of prescribing and adjusting medication. Further study should examine fidelity of the measurements and how to present data for decision-making and feedback.

Tremor may be heightened by many factors. The specific factor (e.g., hypoglycemia) and the tremor state could be used to provide timely information or guidance to the user (e.g., early warning of a hypoglycemic state). Tremor is the kind of subtle input that can take user experience forward into the psychosocial realms of personal experience and interpersonal exchanges. Tremor is an example of a new kind of input that can add to or corroborate other sensory inputs and might be used in reflective- as well as directive-style interfaces.

References

1. Tremor Fact Sheet. National Institute of Neurological Disorders and Stroke, National Institutes of Health, `http://www.ninds.nih.gov/disorders/tremor/detail_tremor.htm`
2. The Causes and Effects of Lack of Sleep, Sleep Deprivation. Sleep-Deprivation.com, June 2 (2010), `http://www.sleep-deprivation.com/articles/types-of-sleep-disorder/lack-of-sleep/index.php`
3. Park, W.: Apple Boosts Global Smartphone Market Share Big-Time, Holds on to No. 3 Position. Into Mobile, May 7 (2010), `http://www.intomobile.com/2010/05/07/apple-boosts-global-Smartphone-market-share-big-time-holds-on-to-no-3-position.html`
4. Smaga, S.: Tremor. American Family Physician 68(8), 1545–1552 (2003)
5. Strachan, S., Murray-Smith, R.: Muscle tremor as an input mechanism. In: UIST 2004 (2004)
6. Zesiewicz, T.A., Hauser, R.A.: Phenomenology and treatment of tremor disorders. Neurol. Clin. 19, 651.vii–680.vii (2001)
7. Maykoski, K.A., Rubin, M.B., Day, A.C.: Effect of Cigarette Smoking on Postural Muscle Tremor. Nursing Research 12, 39–43 (1997)
8. Quek, M.: Background sensing using a mobile device. Diss., University of Glasgow (2008)
9. Buysse, D.J., Reynolds, C., Monk, T., et al.: Pittsburgh Sleep Quality Index: a new instrument for psychiatric practice and research. Psychiatry Res. 28, 193–213 (1989a)

Recognizing Emotions from Video in a Continuous 2D Space

Sergio Ballano[1], Isabelle Hupont[1], Eva Cerezo[2], and Sandra Baldassarri[2]

[1] Aragon Institute of Technology, Department of R&D and Technology Services, Zaragoza. 5018, María de Luna 7-8, Spain
[2] Universidad de Zaragoza, Computer Science and Systems Engineering Department, Zaragoza. 50018, María de Luna 3, Spain
{sballano,ihupont}@ita.es, {ecerezo,sandra}@unizar.es

Abstract. This paper proposes an effective system for continuous facial affect recognition from videos. The system operates in a continuous 2D emotional space, characterized by evaluation and activation factors. It makes use, for each video frame, of a classification method able to output the exact location (2D point coordinates) of a still facial image in that space. It also exploits the Kalman filtering technique to control the 2D point movement along the affective space over time and to improve the robustness of the method by predicting its future locations in cases of temporal facial occlusions or inaccurate tracking.

Keywords: Affective computing, facial expression analysis.

1 Introduction

Facial expressions are often evaluated by classifying still face images into one of the six universal “basic” emotions proposed by Ekman [1] which include “happiness”, “sadness”, “fear”, “anger”, “disgust” and “surprise”. This categorical approach fails to describe the wide range of emotions that occur in daily communication settings and ignores the intensity of emotions.

Given that humans inherently display facial emotions following a continuous temporal pattern [2], more recently attention has been shifted towards sensing facial affect from video sequences. The study of facial expressions’ dynamics reinforces the limitations of categorical approach, since it represents a discrete list of emotions with no real link between them and has no algebra: every emotion must be studied and recognized independently.

This paper proposes a method for continuous facial affect recognition from videos. The system operates in a 2D emotional space, characterized by evaluation and activation factors. It combines a classification method able to output, frame per frame, the exact location (2D point coordinates) of the shown facial image and a Kalman filtering technique that controls the 2D point movement over time through an “emotional kinematics” model. In that way, the system works with a wide range of intermediary affective states and is able to define a continuous emotional path that characterizes the affective video sequence.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 600–603, 2011.

2 Facial Images Classification in a Continuous 2D Affective Space

The starting point of the system is the method for facial emotional classification presented in authors' previous work [3]. The inputs to this method are the variations with respect to the "neutral" face of the set of facial distances and angles shown in Fig. 1. This initial method combines through a majority voting strategy [3] the five most commonly used classifiers in the literature (Multilayer Perceptron, RIPPER, SVM, Naïve Bayes and C4.5) to finally assign at its output a confidence value $CV(E_i)$ of the facial expression to each of Ekman's six emotions plus "neutral".

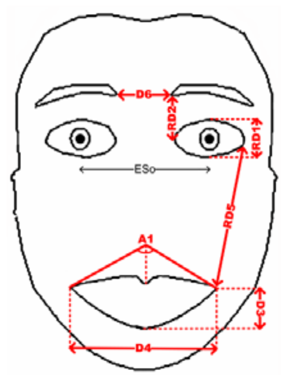


Fig. 1. System's facial inputs

To enrich the emotional output information from the system in terms of intermediate emotions, the evaluation-activation 2D model proposed by Whissell has been used. In her study, Whissell assigns a pair of values <evaluation, activation> to each of the approximately 9000 affective words that make up her "Dictionary of Affect in Language" [4]. The next step is to build an emotional mapping so that an expressional face image can be represented as a point on this plane whose coordinates (x,y) characterize the emotion property of that face.

The words corresponding to each of Ekman's six emotions have a specific location (x_i, y_i) in the Whissell space. Thanks to this, the output of the classifiers (confidence value of the facial expression to each emotional category) can be mapped onto the space. This emotional mapping is carried out considering each of Ekman's six basic emotions plus "neutral" as weighted points in the evaluation-activation space. The weights are assigned depending on the confidence value $CV(E_i)$ obtained for each emotion. The final coordinates (x,y) of a given image are then calculated as the centre of mass of the seven weighted points.

3 From Still Images to Video Sequences

Thanks to the use of the 2-dimensional description of affect, which supports continuous emotional input, an emotional facial video sequence can be viewed as a point (corresponding to the location of a particular affective state in time *t*) moving through this space over time. In that way, the different positions taken by the point (one per frame) and its velocity over time can be related mathematically and modeled, finally obtaining an “emotional path” in the 2D space that reflects intuitively the emotional progress of the user throughout the video.

3.1 Modeling Emotional Kinematics with a Simple Kalman Filter

For real-time “emotional kinematics” control, the Kalman filter is exploited. Analogously to classical mechanics, the “emotional kinematics” of the point in the Whissell space (x-position, y-position, x-velocity and y-velocity) is modeled as the system’s state in the Kalman framework at time t_k. In this way, the Kalman iterative estimation process -that follows the well-known recursive equations detailed in Kalman’s work [5]- can be applied to the recorded user’s emotional video sequence, so that each iteration corresponds to a new video frame (i.e. to a new sample of the computed emotional path). One of the main advantages of using Kalman filter for the 2D point emotional trajectory modeling is that it can be used to tolerate small occlusions or inaccurate tracking so that, when a low level of confidence in the facial tracking is detected, the measurement will not be used and only the filter prediction will be taken as the 2D point position.

3.2 Experimental Results

In order to demonstrate the potential of the proposed “emotional kinematics” model, it has been tested with a set of complex video sequences recorded in an unsupervised setting (VGA webcam quality, different emotions displayed contiguously, facial occlusions, etc.). A total of 15 videos from 3 different users were tested, ranging from 20 to 70 seconds, from which a total of 127 key-frames were extracted to evaluate different key-points of the emotional path.

These key-points were annotated in the Whissell space thanks to 18 volunteers. The collected evaluation data have been used to define a region where each image is considered to be correctly located. The algorithm used to compute the shape of the region is based on Minimum Volume Ellipsoids (MVE) and follows the algorithm described by Kumar and Yildrim [6]. MVE looks for the ellipsoid with the smallest volume that covers a set of data points. The obtained MVEs are used for evaluating results at four different levels, as shown in Table 1. As can be seen, the success rate is 61.90% in the most restrictive case, i.e. with ellipse criteria and rises to 84.92% when considering the activation axis criteria. Finally, Fig. 2 shows an example of emotional path obtained after applying the “emotional kinematics” model.

Table 1. Results obtained in an uncontrolled environment

	Ellipse criteria (success if inside the ellipse)	Quadrant criteria (success if in the same quadrant as the ellipse centre)	Evaluation axis criteria (success if in the same evaluation semi-axis as the ellipse centre)	Activation axis criteria (success if in the same activation semi-axis as the ellipse centre)
Success%	61.90%	74.60%	79.37%	84.92%

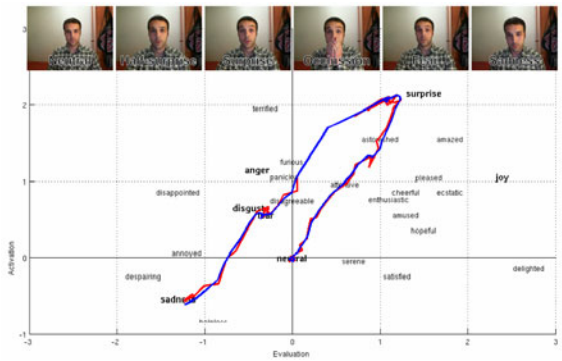


Fig. 2. "Emotional kinematics" model response during the different affective phases of the video and the occlusion period. In dashed red, emotional trajectory without Kalman filtering; In solid blue, reconstructed emotional trajectory using Kalman filter.

4 Conclusions

This paper describes an effective system for continuous facial affect recognition from videos. The main distinguishing feature of our work compared to others is that the output does not simply provide a classification in terms of a set of emotionally discrete labels, but goes further by extending the emotional information over an infinite range of intermediate emotions and by allowing a continuous dynamic emotional trajectory to be detected from complex affective video sequences.

References

1. Keltner, D., Ekman, P.: Facial Expression Of Emotion. In: Handbook of Emotions, pp. 236–249. Guilford Publications, Inc., New York (2000)
2. Petridis, S., Gunes, H., Kaltwang, S., Pantic, M.: Static vs. dynamic modeling of human nonverbal behavior from multiple cues and modalities. In: Proceedings of the 2009 International Conference on Multimodal Interfaces, pp. 23–30 (2009)
3. Hupont, I., Cerezo, E., Baldassarri, S.: Sensing facial emotions in a continuous 2D affective space. Presented at the Systems Man and Cybernetics (SMC), Istanbul, Octubre 10 (2010)
4. Whissell, C.M.: The Dictionary of Affect in Language, Emotion: Theory, Research and Experience. Academic, New York (1989)
5. Kalman, R.: A New Approach to Linear Filtering and Prediction Problems. Transactions of the ASME – Journal of Basic Engineering, 35–45 (1960)
6. Kumar, P., Yildirim, E.A.: Minimum-Volume Enclosing Ellipsoids and Core Sets. Journal of Optimization Theory and Applications 126, 1–21 (2005)

Supporting Moodle-Based Lesson through Visual Analysis

Diego Gomez-Aguilar[1], Miguel Conde-Gonzalez[1],
Roberto Theron[1], and Francisco Garcia-Peñalvo[2]

[1] Computer Science Department
[2] Computer Science Department / Science Education Research Institute /
GRIAL Research Group, University of Salamanca, Spain
{dialgoag,mconde,theron,fgarcia}@usal.es

Abstract. The effective use of CMS requires that instructors can be provided with appropriate means of diagnosing problems. The aim of this research is to support the comprehension of the semantics content evolution within eLearning environments through uncovering by means of visual representations. Therefore, we have carried out the meaning of an eLearning database and represented the more relevant results by depicting them using a visualization based on the tag cloud visual representation. Additionally, we have validated our proposal through a case study.

Keywords: Visualization, e-learning, timeline, tag cloud, Moodle.

1 Introduction

The Moodle is a free license, open-source software platform. The core system of Moodle is structured in modules, each of them providing a wide set of functions. Each module has a connection and access policy based on roles. Also Learning Management System (LMS), or Virtual Learning Environment (VLE) has the same structure. According to [1], Visual Analytics supports analytical reasoning through interactive visual interfaces. This area has an important limitation towards its implementation within the LMS. Using Web Services (WS) in Moodle gives more control in the connecting operation obtaining the optimal service. This API is the base to develop a set of WS: Moodle-DFWSs. This article will use the Moodle web services layer [2] as a proxy to retrieval and exchange information. As the associated data source changes over time, tag clouds can evolve. However, while the tagcloud seems to exposure their evolution over time, they do not explicitly represent them. We organize this paper as follows: in the next section we outline the related work to provide context for our description of Temporal Tagclouds and then the case study, which follows in Section 3. Finally, the contributions and future work.

2 Related Work

A tagcloud usually has the purpose to present a visual overview of a text collection. In the work presented by Torniai et al. the tag cloud employs the size and colour of tags to offer teachers information describing the tags popularity and relevancy[3]. Tree

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 604–607, 2011.

Cloud arranges words on a tree to reflect their semantic proximity according to the text [4]. Despite the significant amount of research on tag clouds, few has been done on how to visualize trends in tag clouds. Parallel Tag Clouds (PTCs) is designed to provide an overview of a document collection by incorporating graphical elements of parallel coordinates with the text size encoding of traditional tag clouds [5]. While PTCs do show multiple clouds simultaneously, they do not explicitly represent trends, and thus comparing multiple tag clouds to ascertain trends places the cognitive demands on the person. Bongshin, et al. integrate sparklines into a tag cloud to convey trends between multiple tag clouds. Also they make controlled study to explore the efficacy of temporal representation on tag cloud, see SparkClouds [6]. The first attempt to use the tag cloud view as a part of the coordinated multiple views (CMV) system was presented by Matkovic, et al. [7].

3 Description of Temporal Tagcloud and Case Study

The main goal of the visualization is to provide a compact representation of the forums' interaction on VLE overall use, and its evolution over the time. Usually, a tagcloud presents a certain number of most often used tags in a defined area of the user interface. A tag's popularity is expressed by its font size (relative to the other tags). Next to their visualization function, tagclouds are also navigation interfaces as the tags are usually hyperlinks leading to a collection of items they are associated with. Furthermore, several layout variations emerged on the basic design principles of tag clouds. Also the representativeness evolution of each tag over the time is one of the goals of the representation. Fig. 1 up provides an overview of a tag with its three main components: the bar-graph, wave-graph representation and the tag. The bar-graph and wave-graph representation curves, shown in the Fig. 1, shows the significance of document content, represented by a word, over time. The size of the words font represents the meaningful of the word on forum activity of the VLE.

Fig. 1. On up,the representation of a single tag on Temporal tagcloud, on down, "view of tag problema"

To further improve readability of the curves and bars meaning, they are assigned with different colours depending on their activity, green for update-post, red for read-post, and blue for add-post. The x-axis encodes the time and the y-axis encodes the significance of the word clouds. Later, this view can be adapted to the user's requirements, so it can explore all the available discussions, forums, courses and users data, going from overview to detail of a given person, course or discussion within a period of time. Here the user can chose, through the contextual menu interaction (see Fig. 2 right), to view among the keywords: the users, the courses, the discussions and the subjects of the forum posts and the obtained from the analysis of all the semantic content of all forum posts exchanged in the platform. One of the interested visual analytic technique used on this work is the semantic zoom [9]. The user can draw or hide every element of the representation on the visualization (notice the difference between Fig. 2 left and Fig. 2 right).

This paper is based upon the study of "view, update and post" statistics obtained through the real time "Reports" from the logs of the Software Engineering course, during the academic year 2008-2009 and 2009-2010, we are considering 88100 log entries. In both courses the subject has been developed in the first semester of the academic year (September to February) and had an average of about 160 students. Regarding the forum analysis, we are going to consider 51 forums, with 114 discussion threads, 172 posts and 26979 forum accesses.

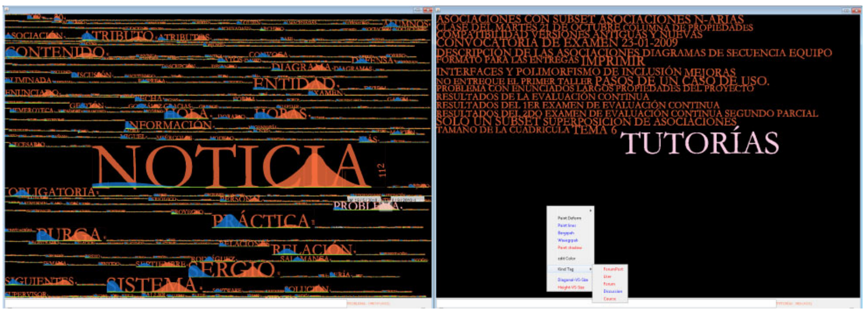


Fig. 2. On left, Global view of temporal tagcloud of forum posts. On right, Global view of temporal tagcloud of discussions.

Using that information we intend to conduct a study which should determine the behaviour of the activity in platform. The result of the application of Temporal Tagcloud is shown in Figure 2 left.

Monitoring a specific theme, forum, discussion or course: the user tries to maintain the track of all discussions, forums and posts that they make, as a normally conversation or discussion. In the Figure 2 left, we could see in a bigger size the most used words . From this figure we could infer that the forum is used to inform students about events related to the subject. Also here must be considered some issues of the workshop statement, such as: Id, Gestión, Sistema, sistema de Purga, etc. Furthermore, the forum is used to answer questions related to different areas of the subject. Also, the users need to review, update and to monitor the most frequent discussions, posts, specific problem or student. Theoretically, the increase should be focused on the months of October and November and may be included a part of December. In Fig. 2 left and Fig. 1 down can be seen that each tagcloud can show (see the "problema" tagcloud) the specific period of time on wave-graph through a mouse move interaction. Therefore, it shows that periods of increased activity in the commented courses correspond to the months of classroom teaching, with particularly representative peaks around the dates of the workshops. In Fig. 1 up, we could observe this phenomenon with the word ¨practice¨. Taking into consideration adding posts (in blue) we could see that the information of the Practice is added at the beginning of the course. Considering update posts (in green) we can see that there are very few updates in Practice information. Regarding the read-post, given its rightward shift of the word, symbolizes that first appears midway through the course and stands towards the end, having a greater frequency near the end of teaching period. The user can select a specific period of the time, through Ctrl+ double click over the period that she/he wants (showed in fig. 1 down).

4 Contributions and Conclusions

Based on our research, this work is one of the first to generate a visualization that uses tag clouds to depict evolving text content over time. As a result, our work offers unique contributions: **The possibility to use some tasks of the tagcloud**: search, browsing, impression formation or gisting, recognition/matching, Understanding evolution of the tag. **Time-based tag cloud layout** that balances semantic coherence of content and spatial stability of the visualization, helping users to perceive easily content updates. Additionally, future improvement could be including new features, like synonyms and plurals binding (with the use of ontology).

References

1. Thomas, J.J., Cook, K.A.: The Science of Analytical Reasoning. In: Illuminating the Path: The Research and Development Agenda for Visual Analytics, pp. 32–68. National Visualization and Analytics Ctr (2005)
2. Conde González, M.Á., García Peñalvo, F.J., Casany Guerrero, M.J., Alier Forment, M.: Adapting LMS architecture to the SOA an Architectural Approach. In: Perry, M., Ehmann, H.S.M., Ortiz Bellot, G., Dini, O. (eds.) Internet and Web Applications and Services, ICIW 2009, pp. 322–327. IEEE Computer Society, Los Alamitos (2009)
3. Torniai, C., Jovanovic, J., Gasevic, D., Bateman, S., Hatala, M.: E-Learning meets the Social Semantic Web. In: Conference E-Learning meets the Social Semantic Web, pp. 389–393 (2008)
4. Gambette, P., Véronis, J.: Visualising a Text with a Tree Cloud. In: Locarek-Junge, H., Weihs, C. (eds.) Classification as a Tool for Research, pp. 561–569. Springer, Heidelberg (2010)
5. Collins, C., Viegas, F.B., Wattenberg, M.: Parallel Tag Clouds to explore and analyze faceted text corpora. In: Conference Parallel Tag Clouds to Explore and Analyze Faceted Text Corpora, pp. 91–98 (2009)
6. Bongshin, L., Riche, N.H., Karlson, A.K., Carpendale, S.: SparkClouds: Visualizing Trends in Tag Clouds. IEEE Transactions on Visualization and Computer Graphics 16, 1182–1189 (2010)
7. Matković, K., Lež, A., Gračanin, D., Ammer, A., Purgathofer, W.: Event line view: Interactive visual analysis of irregular time-dependent data. In: Taylor, R., Boulanger, P., Krüger, A., Olivier, P. (eds.) Smart Graphics. LNCS, vol. 6133, pp. 208–219. Springer, Heidelberg (2010)
8. Weiwei, C., Yingcai, W., Shixia, L., Furu, W., Zhou, M.X., Huamin, Q.: Context preserving dynamic word cloud visualization. In: Proc. Pacificvis 2010, pp. 121–128 (2010)
9. Modjeska, D.: Navigation in Electronic Worlds: Research Review for Depth Oral Exam David Modjeska Department of Computer Science (1997)

Supporting Transformations across User Interface Descriptions at Various Abstraction Levels

Mauro Lisai, Fabio Paternò, Carmen Santoro, and Lucio Davide Spano

CNR-ISTI, HIIS Laboratory,
Via Moruzzi 1, 56124 Pisa, Italy
{mauro.lisai,fabio.paterno,carmen.santoro,
lucio.davide.spano}@isti.cnr.it

Abstract. Model-based approaches for user interfaces exploit various models in order to represent interactive systems at different levels of abstraction. During the design and development process, it is useful to have transformations to derive higher or lower level models. Such transformations should be customizable by designers to reach the desired results. In this paper we present a tool that allows designers without deep knowledge of transformation languages in creating and executing such transformations.

Keywords: Model-based Design, Model-to-model transformation.

1 Introduction

Model-based approaches for User Interfaces (UI) have been applied in many domains, for example to address the problem of creating multi-platform UIs, in which abstract descriptions are transformed first into concrete descriptions and then into various implementation languages. In such approaches, model-to-model (M2M) and model-to-code (M2C) transformations play an important role, as their results should be as close as possible to the designer's expectations. Given that different transformation outputs are desirable for different use cases, the possibility to create and modify transformations is important. Different transformation languages have been proposed in the literature. We focus on those that have been exploited in particular for user interface descriptions. Since many UI specifications are XML-based, it is possible to exploit XSLT (eXtensible Stylesheet Language Transformations) for specifying the transformations. However, even if XSLTs are powerful and supported by many tools, they have a complex syntax that is not really suitable for many UI designer's background. Transformation Templates [1] have been proposed in order to parameterize the transformation logic, according to a set of parameter types. They are relevant for our work and we would like to obtain easier to manage UI transformations representations. In particular, we have considered MARIA [2] (Model-based lAnguage foR Interactive Applications), an XML-based set of model-based languages for which the proposed transformation tool can be applied to define mappings between abstract and concrete levels, as well as the transformations towards implementation languages based on XML.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 608–611, 2011.

2 Tool Support

In this section we describe the interactive environment supporting editing transformations that we propose. Figure 1 shows the UI for the definition of a transformation. Once this part has been activated, the designer can select the desired source and the target meta-model from a drop down list. Each meta-model is represented through a tree view, which allows the designer to easily recognize the hierarchical structure of the composing elements. Figure 1 shows the representation of MARIA desktop CUI (Concrete User Interface) as source and HTML5 as target metamodel (a transformation between these two languages has been specified with this tool). The tree nodes can be collapsed or expanded in order to hide unnecessary nodes, allowing the designer to focus only on the meta-model elements that are needed for defining the current transformation. When a tree node is selected, the corresponding meta-model entity attributes are displayed on the right panel (in Figure 1 the HTML5 body element attributes are visualized on the right panel).

A transformation rule can be defined first selecting an element from the source meta-model (that will be highlighted) and then selecting the destination element from the target meta-model. This action creates a simple transformation rule that maps the source to the target element without any particular condition. An arrow that connects the source and the target nodes represents the existence of such rule.

When the rule has been created, it is possible to add attribute mappings through a dedicated dialog. It shows the list of the currently defined mappings, allowing the designer to create new associations by selecting the source and the target attributes or to remove existing ones. This procedure is sufficient for defining *single* transformations. Moreover, *multiple* transformations can also be defined, iterating the process for creating a single one. Indeed, when more than one arrow starts from a source element, the tool prompts the designer to select the multiple transformation type (conditional, sequential or hierarchical). According to the selected type, the designer can add conditions, ordering and hierarchy values.

A *single* transformation rule can be used for instance in order to define a one-to-one mapping between a MARIA XML *Image* and a HTML *img*. The *sourceImage* attribute can be mapped to the *src* attribute of the *img* element.

A *conditional multiple* transformation rule can be defined adding conditions for the execution of each component. This can be done specifying, through a dedicated dialog, the *selection rule*. The dialog allows the designer to enter a list of conditions between two attributes, or between attributes and a static value. The condition operators that can be selected are different according to the selected attributes data type. An example of such transformation is the MARIA *grouping* mapping towards different HTML5 elements. According to its *role* attribute, a *grouping* can indeed represent one of the following HTML5 elements: *header, nav, section, article, aside* and *footer*. The multiple conditional rule checks the attribute value and generates the corresponding HTML element.

A *sequential multiple* transformation rule needs the specification of the ordering attribute. By default, when the designer selects this type of transformation, the ordering attribute is set according to the single component specification order (the arrow). However, the ordering attribute value is shown on the editor right panel, and it is possible to change it directly. Such rule can be used for instance in order to map a

MARIA *spin box* towards multiple HTML elements. Considering that this element is not included in HTML, the transformation maps it towards a button (with a plus label), a text field *input* element and another button (with a minus label).

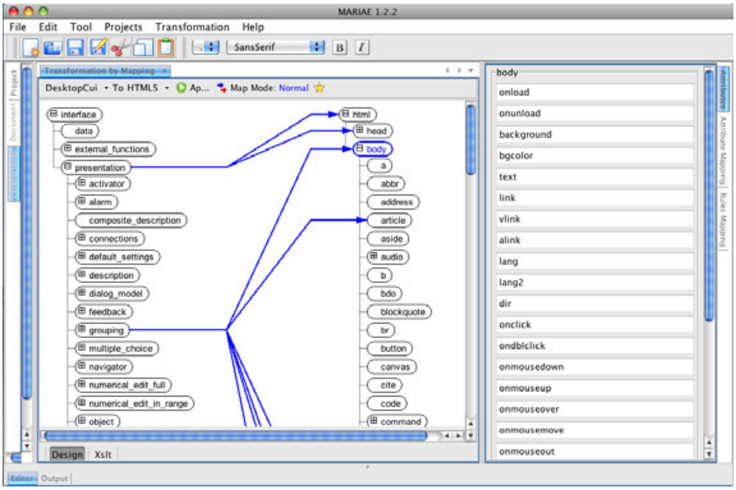


Fig. 1. The Transformation Editor User Interface

For a *hierarchical multiple* transformation, the designer has to specify both the hierarchy and the ordering attribute. By default, the tool creates a new nesting for each component, increasing the current maximum hierarchy value. Default values can be changed through the editor right panel. Such transformation is useful especially in case of model refinement, as it happens in MARIA, where the CUI languages are obtained by adding refinements to the elements of the abstract language. For instance it is possible to refine a MARIA AUI (Abstract User Interface) activator into a desktop CUI button. In this case, multiple elements have to be generated, which should also be nested (the specification of the label is inside the *button* element, which is in turn contained into the *activator* element).

It is also possible to define transformations that map only attributes, without generating target elements. The procedure is the same as before, the designer specifies the different transformation type by pressing a button on the toolbar. Only attribute mappings are displayed using a different colour for the corresponding arrows: blue is used for mappings that generate target elements, red is used with only attribute mappings.

A first user test has been carried out in order to evaluate the effectiveness of the transformation meta-model and the usability of the tool. The test involved 11 users, all male, 27 years old on average. The participants had experience with UI development, and good knowledge of modelling techniques. However, all participants had very little knowledge of transformation languages. Thus, they represent our target users: UI designers without a deep knowledge of transformation techniques.

The participants were required to complete five tasks: loading the source and the target meta-models, editing a single transformation rule, editing a multiple transformation rule, editing an only-attribute mapping, and saving and loading a transformation model. At the end, users were asked to complete a questionnaire, evaluating the transformation meta-model and the tool features on a 1 to 5 scale (with 1 the most negative value and 5 the most positive one). The main results are reported in terms of mean values and standard deviations.

The assessment of arrow-drawing paradigm for creating single transformations was on the positive side (4.45 + 0.69), although some participants expressed some complaints related to some confusion occurring when selecting target elements, due to a lack of knowledge of either the source or the target meta-model. One user proposed introducing a suggestion feature: when the designer is drawing an arrow, the nodes that are semantically similar with the source one should be highlighted with a different colour. Also tool support for creating multiple transformation rules was considered on the positive side (3.73 + 0.90). The participants suggested introducing a brief explanation of the three types of multiple transformations that can be defined with the tool. Moreover, users suggested that it should be possible to identify the type of multiple transformation directly from its graphical representation: the arrow should have a different colour or some other kind of indication. There should also be the possibility to see a summary of a multiple transformation, listing the values of the hierarchy and/or ordering attributes for each component at once. Overall, the evaluation provided encouraging feedback regarding the ways to represent transformations.

3 Conclusions and Acknowledgments

In this paper we discuss a tool for the specification of user interfaces transformations across specifications at various abstraction levels. The tool prototype allows UI designers to create their own transformations and to apply them to the UI meta-models. The tool has been integrated in a model-based authoring environment. The first evaluation of the tool provided positive feedback regarding the transformation meta-model with its tool support. Future work will be dedicated to investigating further refinements in the tool support, together with further empirical evaluation of the approach proposed.

We gratefully thank the support from the EU ICT SERENOA Project (http://www.serenoa-fp7.eu/).

References

1. Aquino, N., Vanderdonckt, J., Pastor, O.: Transformation templates: adding flexibility to model-driven engineering of user interfaces. In: Proceedings of the, ACM Symposium on Applied Computing (SAC 2010), pp. 1195–1202. ACM, New York (2010)
2. Paternò, F., Santoro, C., Spano, L.D.: MARIA: A universal, declarative, multiple abstraction-level language for service-oriented applications in ubiquitous environments. ACM TOCHI 16(4), 19:1–19:30 (2009)

Texture Recognition: Evaluating Force, Vibrotactile and Real Feedback

Jonatan Martínez[1], Arturo S. García[2], Diego Martínez[1], José P. Molina[1], and Pascual González[1]

[1] LoUISE Research Group,
Universidad de Castilla-La Mancha
[2]Symbia IT, Parque Científico y Tecnológico de Albacete
02071 Albacete, Spain
jonatan.martinez@uclm.es, arturo@symbiait.com,
{diegomp1982,jpmolina,pgonzalez}@dsi.uclm.es

Abstract. A force-feedback Phantom device, a custom-built vibrotactile dataglove, and embossed paper sheets are compared to detect different textures. Two types of patterns are used, one formed by different geometrical shapes, and the other with different grooves width. Evaluation shows that the vibrotactile dataglove performs better in the detection of textures where the frequency of tactile stimuli varies, and it is even useful to detect more complex textures.

Keywords: force feedback, vibrotactile; textures, dataglove.

1 Introduction

Haptic feedback is of vital importance in manipulative and exploration tasks of the daily life, as stated in [1]. Unfortunately, many interfaces for virtual environments do not provide this kind of feedback. One of the tasks that have been carried out to verify the effectiveness of this technology is the identification of materials and textures. Minsky et al. [2] used a joystick to experiment with force feedback, using a depth map texture. Some other authors attempted to optimize these systems when used to distinguish different materials [3] [4]. Tactile feedback can be used as a complement to force feedback, but it is also useful by its own, and it is often used even to replace it. In particular, vibrotactile feedback uses vibrations to transmit sensations through the skin, which plays an essential role in the way the different textures are detected [5]. There are tasks in which vibrations can increase the performance, reducing the response times or minimizing the forces used. Small modified speakers were used in [6] and, in a similar experiment, [7] used vibrotactile tactors to discriminate materials of different stiffness. A different actuator, the vibrator motor, has reduced bandwidth but is integrated in multiple devices, such as gaming peripherals [8], mobile phones, and datagloves [9].

In this context of texture identification, Kyung et al. [10] describes an experiment that compares force, tactile and vibrotactile feedback technologies. It was considered of interest to continue this experiment and use the same basis in order to have a reference to compare the results with. In the proposed study, some changes have been

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 612–615, 2011.

introduced, the vibrator is integrated in a dataglove and located directly on the fingertip, one of the most sensitive areas of the body [11], the control algorithm has been be optimized to reduce its latency, and some paper patterns have been introduced so that a real model can be considered in the analysis of the results. Next section will detail the patterns and the haptic feedback methods used.

2 Description of the Haptic Feedback Methods

The comparison of the three haptic feedback methods was performed using a discrimination task involving the identification of textures that followed some predefined patterns. Two of the three groups of patterns shown in [10] were used in this experiment. In the first group, each pattern is composed of four times the same geometric shape, while in the second each pattern is formed by horizontal lines with different spacing between them. These patterns were converted into tangible images where the black areas are 1 mm deeper than white ones (Fig 1 -left). In order to distinguish them, the user must perform scanning movements with each of the haptic approaches considered:

a) **Force feedback - Sensable Phantom.** With this device [12] and the H3DAPI library [13], textures were represented in a virtual box with patterns embossed in its upper side, using a depth map where the gray level determined the relative displacement. An X3D model was created for each pattern, and the Ruspini algorithm was selected for the haptic rendering.
b) **Vibrotactile feedback.** The second method is a dataglove capable of providing vibrotactile feedback that was developed in our laboratory [14]. This device has a small vibrating actuator, located on the index finger, whose frequency can be varied by the pulse width modulation generated on a microcontroller. The finger was tracked by a PhaseSpace system, attaching one LED on top of the fingertip (Fig. 1 –right).
c) **Direct stimulation.** In this case, the user moves her finger directly on the texture. It is the ideal tactile feedback, because the latency is zero, and the bandwidth and resolution are only limited by the sensitivity of the skin. The patterns are built using transparent paper 1 mm thick, removing the black areas to create zones of palpable depression. Patterns were placed on a table beneath another larger table that hid it form the user.

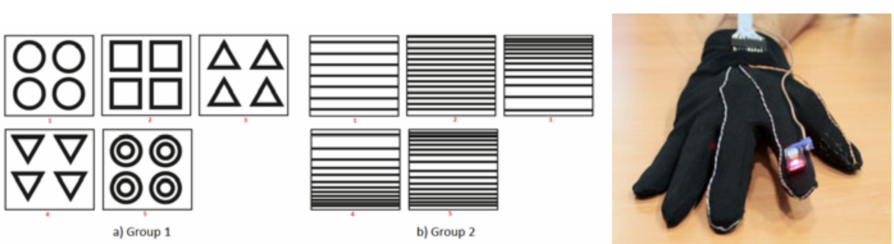


Fig. 1. Groups of patterns used (left) and custom-made dataglove (right)

3 Experiment Design

Twelve different users (4 women and 8 men) were requested to distinguish the patterns of both groups using the three aforementioned methods. After each trial the users were informed about the correct answer. To prevent that the order could influence the results, the sequence was determined by the Latin square method. After each test, users were asked to rate the time they spent learning to detect patterns, the difficulty to distinguish them in an advanced phase, and the comfort of each device.

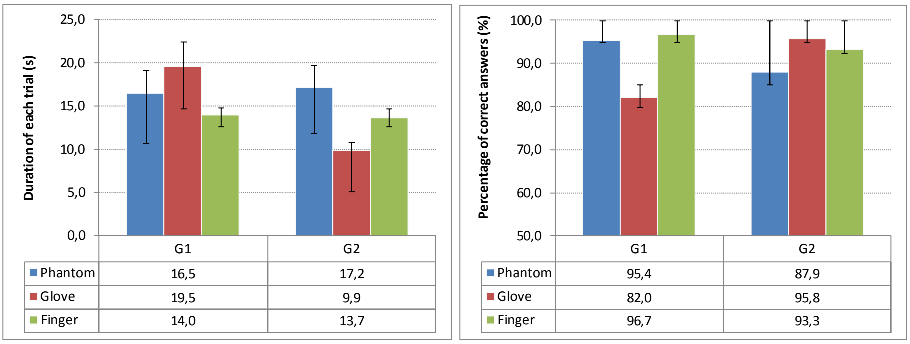


Fig. 2. Average duration of trials (left) and percentage of correct answers (right). Bounded lines represent the interval between the first and third quartile of the samples.

4 Results and Discussion

Fig. 2 shows both the average duration of each attempt (left), and the average percentage of correct answers (right). As expected for the first group of textures, the method of direct tactile stimulation is the fastest and the one which provides the higher percentage of correct answers, due to the advantage of having the entire surface of the fingertip to follow contours and identify shapes. Similarly, the force feedback percentage of hits is very close, because the device guides the user's finger when the cursor passes over an area of depression, reducing the cognitive effort. In the case of the dataglove, the stimulation is performed in one point and does not allow accurate tracking of the border. In this case, the user is required to develop a detection strategy different from the one followed naturally. This requires the user to make a greater effort, resulting in higher error rate and time consumption.

For the second set of textures, the most efficient method in terms of both time and hit rate is the vibrotactile feedback, even improving the method of direct stimulation. To discriminate between different patterns, the user typically scans the texture across the lines at constant speed, trying to identify the timing or frequency of the marks. That is why the vibration is appropriate, since the user perceives clearly the necessary information. In contrast, when the user swipes its finger across the paper textures much more spatial information is received that has to be discarded, so the effectiveness is not as good in terms of time and error rate. Finally, the separation between the lines is perceived mostly at a kinaesthetic level when using the force feedback, affecting the performance of this feedback method in this test.

5 Conclusions

A texture discrimination experiment has been conducted in order to compare force and tactile feedback with the vibration feedback of a glove developed at our laboratory. Vibrotactile feedback seems to be the most effective method to distinguish between texture patterns that can be identified by the frequency changes of their surface features while rubbing it with the finger. However, in tasks where a precise spatial recognition is needed to identify shapes it has not resulted as effective as other methods, yet has proved to be useful.

This study shows some interesting results that can be corroborated in a future work by expanding both the number of users and the variety of textures to detect, also recognizing textures or shapes in the space, where the use of multiple vibrotactile actuators can be an advantage compared to the localized force feedback.

Acknowledgments. This work has been supported by the projects PEII09-0054-9581 and TIN2008-06596-C02-01. Thanks to the users who have participated voluntarily in the experiments.

References

1. Robles-De-La-Torre, G.: The Importance of the Sense of Touch in Virtual and Real Environments. IEEE Multimedia, 24–30 (2006)
2. Minsky, M., Ming, O., Steele, O., Brooks Jr, F.P., Behensky, M.: Feeling and seeing: issues in force display. In: ACM Symposium on Interactive 3D graphics, pp. 235–241 (1990)
3. Kuchenbecker, K., Fiene, J., Niemeyer, G.: Improving contact realism through event-based haptic feedback. IEEE Transactions on Visualization (2006)
4. Okamura, A.K., Dennerlein, J., Howe, R.D.: Vibration feedback models for virtual environments. In: IEEE Int. Conf. on Telerobotics and Automation, pp. 1–6 (1998)
5. Hollins, M., Bensmaïa, S., Roy, E.: Vibrotaction and texture perception. Behavioural Brain Research 135, 51–56 (2002)
6. Wellman, P., Howe, R.D.: Towards realistic vibrotactile display in virtual environments. In: Symp. on Haptic Interfaces for Virtual Env. and Teleoperator Sys. (1995)
7. Gurari, N., Smith, K., Madhav, M., Okamura, A.M.: Environment discrimination with vibration feedback to the foot, arm, and fingertip. In: IEEE Int. Conf. on Rehabilitation Robotics, pp. 343–348 (June 2009)
8. Nintendo Rumble Pak, `http://www.nintendo.es/NOE/es_ES/systems/accesorios_1225.html`
9. CyberGlove Systems, `http://www.cyberglovesystems.com/`
10. Kyung, K.U., Lee, J.Y., Park, J.S.: Comparison of force, tactile and vibrotactile feedback for texture representation using a combined haptic feedback interface. In: Oakley, I., Brewster, S. (eds.) HAID 2007. LNCS, vol. 4813, pp. 34–43. Springer, Heidelberg (2007)
11. Johansson, R.S.: Tactile Sensibility in the human hand: receptive field characteristics of mechanoreceptive units in the glabrous skin area. Journal of Physiol. 281 (1978)
12. Sensable, `http://www.sensable.com`
13. H3D Open Source Haptics, `http://www.h3dapi.org`
14. Martínez, J., Molina, J.P., García, A.S., Martínez, D., González, P.: Desarrollo de un Guante de Datos con Retorno Háptico Vibro-táctil Basado en Arduino. JoReVir, 1–10 (2009)

The Application of Preference Mapping in Aesthetic Website Evaluation

Eleftherios Papachristos and Nikolaos Avouris

Human-Computer Interaction Group, Electrical and Computer Eng. Dept.,
University of Patras, GR-265 00 Rio Patras, Greece
{epap,avouris}@ece.upatras.gr

Abstract. The objective of this study was to apply a technique called preference mapping to the context of aesthetic website evaluation. Preference mapping is a method in which evaluators and stimuli are simultaneously represented in the same multidimensional space. User segmentations and drivers of preference can easily be identified. We argue that this technique is particularly suited for website design evaluation especially for alternative prototype comparisons. The application of this method to an actual dataset resulted in a better understanding of participant preferences that could not be reached through simple comparison of average ratings.

Keywords: Web design, aesthetic evaluation, preference mapping.

1 Introduction

The importance of aesthetics in web design has been demonstrated by several studies [1, 2]. This realization underlines the need for visual website evaluation methodologies during initial stages of development. In this paper we tried to demonstrate that preference mapping is an appropriate method for the evaluation of website aesthetics. Preference mapping is referred to as a group of multivariate statistical techniques aimed at gaining deeper understanding of preference toward products by taking also participants' heterogeneity into account [3]. It is not a new technique since it is used widely in sensory analysis and marketing for many years. The main goal is to find relationships between descriptive data (usually provided by domain experts) and hedonic judgments of evaluators. This methodology can be encountered in HCI literature or website evaluation only rarely with the exception of [1].

There are two main preference mapping techniques which are generally referred to as internal (IPM) and external (EPM) analysis. There is as ongoing debate about the advantages and disadvantages of these variations of preference mapping [4]. In this study we used IPM alongside hierarchical cluster analysis to identify participant preference segmentation. The goal of IPM is to derive a multidimensional representation of objects and evaluators. This is usually accomplished by conducting principal component analysis (PCA) on a data matrix consisting of objects in rows and participants (treated as variables) in columns. In a variation of the technique called extended IPM, external descriptive data (usually provided by expert panel) are projected onto the preference space. This can be done by projecting the attribute mean scores onto the map coordinates through regression [3].

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 616–619, 2011.

2 Method

The dataset used was previously acquired by a study regarding subjective and objective factors that influence website design evaluation [5]. In this study, 53 volunteer students (35male, 18 female, mean age = 22.2) had to choose a new website design for their university department among six prototypes. The test designs were evaluated by two sources: the volunteer students and a trained panel of 9 design experts. Student participants first viewed all the website screenshots in a random order and then rated them on *visual appeal*, *credibility, perceived usability*, *novelty* and *overall preference* on a linear, unmarked scale (from 0 to 100). The trained panel first helped in generating a list of 14 design related descriptive attributes and then rated the test websites on them.

3 Results

PCA was applied to participant's preference data for the websites. A two factor solution explained 68.6% of the variance in participant preference (Fig. 1).

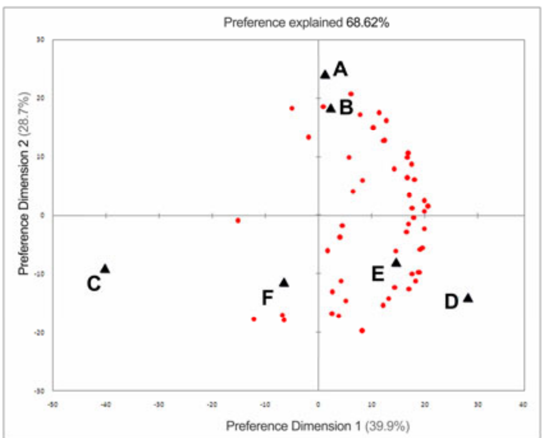


Fig. 1. Bi-plot of websites and individual user preference in the derived two-dimensional space

Designs that are close together were perceived as similar regarding preference. Each dot represents an individual user. A line from the center of the axes to that dot shows general direction of preference for that participant. If all users were pointing to approximately the same direction then their preferences would be homogenous and no segmentation would be necessary. Although almost all preference is pointing to the right, there is a significant spread across the vertical axis and sample segmentation seemed required. Hierarchical cluster analysis on preferences revealed 3 groups of participants (table 1) with very different preferences. Group 1 gave moderate to high ratings to most websites except of the first two. Group preferred these first two designs the most and clearly disliked designs C and F. Finally, Group 2 which was also the most populated, was closer to the overall preference ratings, with clear preference to the last 3 designs and extreme dislike of design C.

Table 1. Mean design preference scores for the whole sample and the identified subgroups

	A	B	C	D	E	F
All users (n=53)	50,9	48,6	24,9	68,9	60,5	50,4
Group 1 (n=15)	38,3	39,6	56,6	63,1	50,9	66,6
Group 2 (n=22)	39,7	44,8	13,3	84,6	77,8	55,4
Group 3 (n=16)	78,1	62,2	11,3	52,7	45,9	28,5

In order to understand the drivers of preferences for our sample as a whole and for the different subgroups the derived preference dimensions have to be interpreted. This can be accomplished either by visual inspection of the websites and their position in the map or by projection of external descriptive data. The first method is more subjective but extremely useful if no external descriptive data is available. The projection of descriptive data requires the regression of their mean ratings in the dimensional space. We projected the evaluative constructs from the participants and the descriptive attributes from our expert panel that fitted significantly. The result of the attribute projection is depicted in Figure 2.

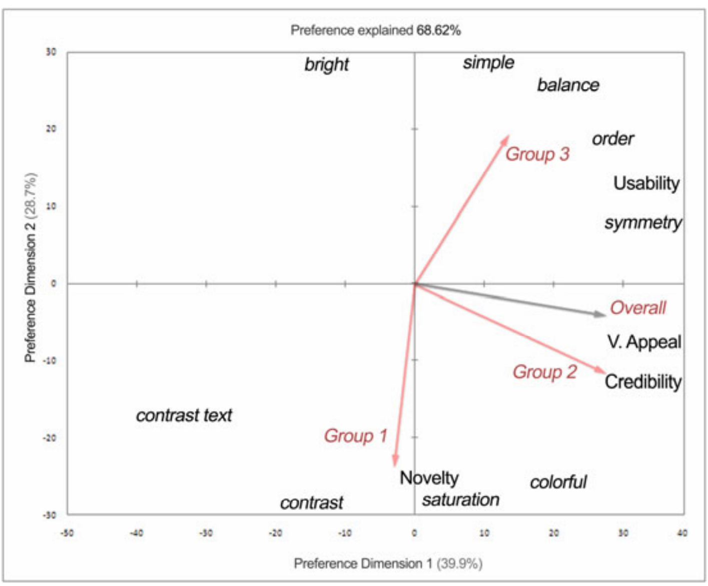


Fig. 2. Attribute projection in the preference space. Arrows represent direction of preference for the user clusters.

Examination of the final IPM map revealed that the overall preference direction and therefore the main driver for the participants as a whole, was *visual appeal*. Group 1 seems to be more influenced by *novelty*, group 2 by *credibility* and group 3 by *perceived usability*. Constructs that are close to each other, such us *credibility* and *visual appeal*, are highly correlated. Analysis of descriptive attribute positions reveals that increased *order*, *balance* and *simplicity* results in more *perceived usability* but also decreases perceptions of *visual appeal* and *novelty*. *Novelty* perceptions are driven by increased *saturation*, *contrast* and *colorful*. *Visual appeal* and *credibility* seem to require balance on the above attributes. The preference direction of overall preference and of the most populated group is towards designs E and D. The examination of table 1 reveals that design D is highly rated by each group and therefore is the safest pick as the final design choice.

4 Discussion

Our goal was to demonstrate the advantages of preference mapping in the context of visual web design evaluation by alternative prototype choice. The application of this technique resulted in deeper understanding of user perceptions and their key drivers of preference. The methodology helped us make better informed decision about final choice and also gave us instructions for design improvement towards desired directions. In addition the process of user clustering regarding their preference revealed distinct groups of participants with different opinions about final choice.

However, the above results cannot be generalized and are specific to the test sample of our case study. Our goal was to demonstrate an evaluation methodology which could be tailored to the context of different studies. To conclude, preference mapping seems promising in the visual web design evaluation context and flexible enough to be used in several different conditions.

References

1. Schenkman, B.N., Jonsson, F.U.: Aesthetics and preferences of web pages. Behavior and Information Technology 19(5), 367–377 (2000)
2. Tractinsky, N., Katz, A.S., Ikar, D.: What is beautiful is usable. Interacting with Computers 13, 127–145 (2000)
3. Meullenet, J., Xiong, R., Findlay, C., Knovel: Multivariate and probabilistic analysis of sensory science problems. Blackwell Publishing, USA (2007)
4. van Kleef, E., van Trijp, H.C.M., Luning, P.: Internal versus external preference analysis: An exploratory study on end-user evaluation. Food Quality and Preference 17, 387–399 (2006)
5. Papachristos, E., Avouris, N.: The Subjective and Objective Nature of Website Aesthetic Impressions. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5726, pp. 119–122. Springer, Heidelberg (2009)

The Effect of Religious Identity on User Judgment of Website Quality

Ons Al-shamaileh[1], Alistair Sutcliffe[1], and Antonella De Angeli[2]

[1] Manchester Business School,
University of Manchester UK
[2] University of Trento Italy
Alshamaileh.ons@gmail.com,
alistair.sutcliffe@mbs.ac.uk,
antonella.deangeli@disi.unitn.it

Abstract. The paper investigates the effect of users' religious identity on their judgments of website quality. Websites related to Islamic and Christian identities were evaluated by Christian and Muslim respondents. Aesthetics, usability, service quality, pleasurable interaction, content, website identity and overall judgment were assessed, showing that respondents were more positive to the website which related to their own belief but the effect was stronger with the Muslim sample. Interviews were conducted to support the above results with a non-religious well known branded website added. Respondents provided consistent evaluations preferring the website matching their beliefs but brand identity showed to be more important than the religious identity.

Keywords: User experience, user judgment, Religion, brand.

1 Introduction

A conceptual framework for user experience has suggested that the interaction with a product gives rise to emotions, attitudes and thoughts which will influence intentions and interactions with the product used over time; claiming that user experience is affected by individual differences such as knowledge, concerns, skills, personality and physical attributes [2]. Factors such as product, user, context of use and social and cultural factors are also claimed to affect the user experience [1]. To influence users' overall judgment, Organization brand has been shown in market related studies to have a positive effect of favorable brand on quality perceptions [5]. Research has covered several user characteristics and background aspects affecting website overall judgment. However, user religion as a factor affecting user experience has received little attention by researchers; further work on user religious identity is needed. This paper explores user social identity (religion) as one of the factors affecting user judgment of website quality[1]

[1] A composite of the factors measured in this study (aesthetics, usability, service quality, pleasurable interaction, content, website identity and overall judgment).

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 620–623, 2011.

2 Method

2.1 Online Questionnaires

Questionnaires explored the role of religion on user judgment of website quality. A sample of six websites from three different areas (health www.muslimhealthnetwork.org/, www.christianhospital.org/, books, www.islamicbookstore.com, www.christianbook.com and aid, www.muslimaid.org, www.christianaid.org/) associated to Muslim and Christian identities were randomly allocated to users. Respondents' viewed one website and evaluated it by completing an online questionnaire. 371 respondents took part in the online questionnaire, 167 respondents were Muslims; 147 were Christians, 44 atheists, 10 from other religion and three refused to declare their religion. 52% of the sample was aged 18-25, 36% aged 26-35, 7% 36-45 and 5% were older. Only responses from Muslim and Christian respondents were considered in the analysis of this study. The questionnaire was divided into three parts, website evaluation, religious commitment [10] and demographic information. The first part of the questionnaire consisted of several scales; classical aesthetics [7], expressive aesthetics [7], usability [7], pleasurable interaction [7], service quality [7], information quality-content with a nine items was originally adapted from Bernier Instructional Design [3] symbolism [9], overall judgment and behavioral intentions. The total number of questions was 54 questions.

2.2 Interviews

Interviews were conducted to get a deeper insight of the effect of the website religious identity on respondents' evaluations. Brand was included as an additional factor in the interview study; to further examine the difference between social identity (religion) and brand. Twenty eight randomly selected respondents were interviewed. Fifteen were Muslims and thirteen were Christians, age range 22-34. Respondents viewed three websites; two were used in the survey (www.islamicbookstore.com and www.christianbook.com) with addition of a third -non religious- well known brand (http://bookshop.blackwell.co.uk). Participants were then interviewed to express their preference towards each website, and to indicate the most important features affected their decision. Interviews transcripts were coded to summarize respondents' opinions.

3 Results

Cronbach's alpha values ranged from 0.818- to 0.961 for all scales. Results showed that religious identity is a strong influence. Two-way ANOVAs; (respondents religions* websites identities) results showed that website religion had a main effect on user judgment of aesthetics, usability, service quality, pleasurable interaction, content, website identity and overall judgment, with significant interactions of the website identity and the respondent religion for all variables except aesthetics and usability. One way ANOVAs were conducted to analyze differences between Muslim and Christian websites. Results showed that there was significant difference between Muslims and Christians evaluating Muslim websites in all dependent variables except classic aesthetic and usability; while only three factors; pleasurable interaction, symbolism and service quality were significant in evaluating Christian websites.

The Christian sample showed a bimodal distribution and was divided into two groups; strong and weak believers. Using a median split; 84< 5 and were categorized as weak believers. 63 >5 and were categorized as strong believers. One way ANOVAs of the Christian respondents (strong and weak believers) showed no differences while evaluating Muslim websites. However, all scales apart from classic aesthetics and usability were significant while evaluating Christian websites. Strong Christian believers rated Christian websites higher than weak Christian believers, while the strength of religious beliefs did not have an effect when evaluating websites from a non matching identity. Respondents were asked to rank criteria according to their importance in influencing their overall judgments. Content was the most important criterion in all evaluations apart from Muslims evaluating the Christian website where usability was ranked number one and the neutral (branded) website where ID was ranked as the most important criterion. Usability was highly ranked while aesthetics and pleasurable interaction were least important.

The interview results supported the survey indicating that respondents' prefer matching identity websites. Respondents explicitly mentioned that website identity and content are the reasons of preferring the matching website rather than design quality; (Christian on a Muslim website) "everything in the website looks neutral to me apart from the content and religious identity which I consider as negative since it's not of my interest". Similarly; a Muslim on the Christian website; "I will not visit it again; good in terms of design not content". A curiosity effect appeared for Christians expressing interest in Islam as a new experience. "*It is something new and everything new you become curious to see what it is about*". A possible social sub identity emerged in the Christian sample where three respondents did not like the content of the matching identity website because it followed an approach that did not match their beliefs; e.g. "*The content of the website is pushy and not really related to my beliefs, though I am a Christian I feel the content approach is American which is different from mine*". Brand showed a strong effect on users' evaluations, 9/13 Christians and 12/16 Muslims favored branded website over websites with religious identity. Respondents were asked to assess each criterion by indicating (positive, negative or neutral). The net ratings of criteria excluding neutral evaluations were calculated. In matching identity websites; respondents were more positive rating content and website identity in matching than in non matching identity websites. But overall; Muslims and Christians preferred the identity of the branded website which they were all familiar with.

4 Discussion

Results showed that religion as a form of user social identity had an important effect on user judgment of website quality, which agrees with Hartmann et al. [6] indicating that users' background influence the decision making process. This paper showed that websites with religious identity affected users' perception of service quality, pleasurable interaction, content, overall judgment but not usability and aesthetics. People may be separating evaluation of usability and aesthetics from identity and treating them objectively. Respondents favored matching identity websites and evaluated them more favorably than websites from other religions (both Muslims and Christians); which partially agrees with Siala et al [8] results that only Muslims trusted matching identity websites while Christians favored neutral websites. Brand affected respondents' judgment of content, service quality, usability and aesthetics which agrees with De Angeli et al [4] who demonstrated that people who have positive attitudes towards the website brand were more positive in the evaluation of aesthetics, pleasure and usability. This paper demonstrated clear effects of religion on user judgment of website quality with regards to the websites that has a religious identity regardless of its domain; however, some limitations should be noted. The study consisted of several measures representing opinions; these opinions may differ with intentions or actual actions i.e. respondents may have different judgment while only viewing a charity website and donating to that website.

Refrences

1. Arhippainen, L., Tähti, M.: Empirical Evaluation of User Experience in Two Adaptive Mobile Application Prototypes. In: Proceedings of the 2nd International Conference on Mobile and Ubiquitous Multimedia, pp. 27–34. ACM, Norrköping (2003)
2. Beauregard, R., Corriveau, P.: User Experience Quality: A Conceptual Framework for Goal Setting and Measurement. In: Duffy, V.G. (ed.) Digital Human Modeling. LNCS, pp. 325–332 (2007)
3. De Angeli, A., Sutcliffe, A., Hartmann, J.: Interaction, usability and aesthetics: what influences users' preferences? In: DIS 2006: Proceedings of the 6th Conference on Designing Interactive Systems, pp. 271–280. ACM, New York (2006)
4. De Angeli, A., Hartmann, J., Sutcliffe, A.G.: The Effect of Brand on the Evaluation of Websites. Interact (2), 638–651 (2009)
5. Dodds, W.B., Monroe, K.B., Grewal, D.: Effects of Price, Brand and Store Information on Buyers' Product Evaluations. Journal of Marketing Research 28, 307–319 (1991)
6. Hartmann, J., Sutcliffe, A.G., De Angeli, A.: Towards a Theory of User Judgment of Aesthetics and user Interface Quality. ACM Transactions on Computer-Human Interaction (TOCHI) 15(4), Article No. 15 (2008)
7. Lavie, T., Tractinsky, N.: Assessing Dimensions of Perceived Visual Aesthetics of Web Sites. International Journal of Human-Computer Studies 60(3), 269–298 (2004)
8. Siala, H., O'Keefe, R.M., Hone, K.S.: The impact of religious affiliation on trust in the context of electronic commerce. Interacting with Computers 16, 7–27 (2004)
9. Tractinsky, N., Zmiri, D.: Exploring Attributes of Skins as Potential Antecedents of Emotion in HCI. In: Fishwick, P. (ed.) Aesthetic Computing, MIT Press, Cambridge (2005) (in press)
10. Worthington Jr., E.L., Wade, N.G., Hight, T.L., McCullough, M.E., Berry, J.T., Ripley, J.S., Berry, J.W., Schmitt, M.M., Bursley, K.H.: The religious commitment inventory-10: development, refinement and validation of a brief scale for research and counseling. Journal of Counselling Psychology 50(1), 84–96 (2003)

Toward a Better Guidance in Wearable Electronic Orientation Aids

Slim Kammoun, Marc J.-M. Macé, Bernard Oriola, and Christophe Jouffrais

IRIT, CNRS & University of Toulouse,
Toulouse, France
{Kammoun,Mace,Oriola,Jouffrais}@irit.fr

Abstract. Electronic Orientation Aids (EOA) usually guide visually impaired pedestrians using turn-by-turn instructions. However, several studies have demonstrated that mental representations of space seem difficult to build when turn-by-turn instructions only are provided. This is a crucial point as getting an accurate spatial representation of the surroundings can make the difference between successful and unsuccessful navigation. In this paper, we describe components of a guidance system designed to provide a better service in wearable electronic orientation aids for the visually impaired. They especially include non-visual landmarks and points of interest that are used as environmental features to improve mental spatial representations.

Keywords: Visually impaired, Electronic Orientation Aids, navigation, orientation, mobility, assistive technology, guidance.

1 Introduction

In assisted navigation for the pedestrian, guidance instructions usually rely on turn-by-turn instructions (e.g. "turn left in fifty meters"). For visually impaired as well as sighted pedestrians, landmarks play an important role in decision-making when traveling. They are used either as reorientation or confirmation points, and are significant elements in the communication of route directions, which shows that they are useful to elaborate mental representations of space. Hence providing the visually impaired with these environmental features during the guidance process is critical for the usability of EOAs, and therefore for successful navigation tasks, with the parallel benefit of contributing to the elaboration of mental representation of space.

In this paper we defined primary steps toward a better guidance process by taking into account the needs of the visually impaired and the potential of technologies such as wearable sensors, Global Positioning System (GPS), and Geographic Information System (GIS). Guidance consists first in identifying the location of the VI user relative to the expected trajectory, and then in providing him/her with appropriate direction instructions and pertinent information about the surroundings. Hence we defined guidance as three main steps: 1) Route selection to compute the optimal itinerary, 2) Tracking to estimate the location of the traveler, 3) Display of navigation instructions and selected spatial cues to guide the traveler and improve his mental representation of the environment.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 624–627, 2011.

2 Guidance in Electronic Orientation Aids

In the following section, we decompose and describe the guidance process generally used in EOAs, with a particular focus on the differences we propose compared to existing systems.

2.1 GIS and Route Selection

Route selection is defined as the procedure of choosing an optimal path between origin and destination. Based on multiple brainstorming and interviews with VI people and orientation and mobility (O&M) instructors, Kammoun et al. [1] proposed an annotation of geographical data that takes into account the needs of VI pedestrians. This classification included three main classes: 1) The pedestrian network (sidewalks, pedestrian crossing, etc.); 2) Non-visual landmarks (decision or confirmation points) corresponding to locations that can be detected by the VI pedestrian (e.g. pavement texture or inclination, street furniture, specific odor/sound, etc.); 3) Points of Interest (POIs) such as places or objects that are potential destinations (e.g. public buildings, shops, etc.). Landmarks and POIs are also useful clues to construct a more complete mental representation of the environment while traveling. Using such annotations, it is possible to select the route most adapted to the needs of VI pedestrians [1].

2.2 Tracking User Location

Geolocation based on GPS is the most common technique used in EOAs. However, positioning precision with GPS alone is rarely better than 10 to 20 meters in many environments, mostly in cities where urban canyons prevent direct line of sight with the satellites. Several approaches have been developed to improve GPS precision. These solutions include dead-reckoning algorithms, differential GPS (D-GPS) or electronic location identifiers, such as RFID tags, WLAN networks or Bluetooth beacons which are widely used in indoor environments. In order to get a better pedestrian positioning, it is possible to fuse vision-based (local) and GPS-based (global) positioning signals. The EOA described in the following sections included a GPS and embedded stereo cameras. An ultra-rapid recognition and localization algorithm detects geolocated visual landmarks in the environment. The location of the pedestrian relative to the geolocated visual landmarks is then used to refine the global position of the pedestrian on the digitized map.

2.3 Rendering of Descriptive and Spatial Information

Interviews of VI users and O&M instructors made clear that EOAs should provide both directional instructions and information about the surroundings. To display turn-by-turn instructions and space-related information to the VI traveler, an adapted interface is required, which must rely on non-visual (e.g. auditory or somatosensory) modalities. Evaluation showed that both TTS (see e.g. [2] [3]) and spatialized sounds [4] may provide efficient rendering. While TTS technique is perfectly adapted to turn-by-turn instruction, binaural synthesis can be used to generate sounds at any given spatial location relative to the listener. Combination of TTS and binaural synthesis is an interesting solution to display both guidance instructions and space configuration information. Another interesting wearable device consists of actuators mounted on a waist belt [5], but it provides information regarding heading direction only.

3 The NAVIG Orientation Aid

The NAVIG system (Navigation Assisted by artificial VIsion and GNSS) is an assistive device whose objective is to increase the autonomy of visually impaired users in known and unknown environments [6]. Here, we focus on the guidance process that we have implemented in the prototype. Positioning data was provided by a GPS receiver combined with a wearable inertial measurement unit (IMU) and stereo cameras [7]. The IMU included accelerometers, an electronic compass and a pedometer and was developed by the firm and project partner Navocap. The stereo cameras allowed to estimate the traveler's location (direction and distance) relative to environmental features that were geolocated and tagged in the GIS database. This vision-based localization module was developed by the firm and project partner SpikeNet. The GIS module consisted in an adapted digitized map where walking areas (e.g. sidewalks, pedestrian crossings, etc.) and environmental features (visual points for the cameras as well as non-visual landmarks and points of interest for the user) were added to additional layers. Interaction with the user was based on speech recognition. For output interaction, we used a real-time binaural synthesis module (designed by the project partner LIMSI) to render directions, and a TTS module to describe surrounding landmarks and POIs.

When a route was selected between two points, several sections defined by two successive geolocalized Itinerary Points (IPs) were generated. Each section also contained a list of geolocalized landmarks and POIs extracted from the GIS. Using IPs, the system then generated turn-by-turn instructions based on the traveler position and direction provided by the positioning component. Both landmarks and POIs were displayed to provide the user with information about her/his local environment. To track the user location, we used a simple algorithm based on activation fields. To trigger the display of information, a radius was defined according to each type of point. The radiuses of the different activation fields were determined during preliminary tests that were conducted with the positioning module. TTS and Spatialized TTS were used to display environmental information when reaching activation fields.

4 Conclusion and Future Work

In this paper, we focused on the general guidance process designed for an Electronic Orientation Aid for the visually impaired. This process was decomposed in three important components: (1) route selection procedure to compute the optimal itinerary, (2) user tracking to estimate the location of the traveler, and (3) improvement of the mental representation of space through the display of adapted instructions and selected surrounding elements. We implemented this function in a prototype of the NAVIG system, a wearable orientation aid designed to overcome blind mobility difficulties. To be effective, this guidance process relies on an adapted GIS, an improved positioning module (compared to GPS alone), and an adapted user interface based on spatialized sounds and TTS. Further work is necessary to evaluate the impact of this guidance process on navigation tasks. In addition, as mental representations of space are a key factor in the successful accomplishment of navigation tasks, it will also be of interest to evaluate the quality of mental representation built through this guidance process.

Acknowledgments. This work was supported by the French National Research Agency (ANR) through TecSan program (project NAVIG ANR-08TECS-011) and the Midi-Pyrénées region through APRRTT program. The NAVIG consortium includes three research labs (IRIT, LIMSI, and CerCo), two companies (SpikeNet Technology, and NAVOCAP), the Institute for Young Blind of Toulouse (CESDV-IJA), and the community of Grand Toulouse.

References

1. Kammoun, S., Dramas, F., Oriola, B., Jouffrais, C.: Route selection algorithm for Blind pedestrian. In: Proceedings of the International Conference on Control Automation and Systems (ICCAS), pp. 2223–2228. IEEE, Los Alamitos (2010)
2. Gaunet, F.: Verbal guidance rules for a localized wayfinding aid intended for blind pedestrians in urban areas. Universal Access in the Information Society 4, 338–353 (2006)
3. Golledge, R., Marston, J., Costanzo, C.M.: Assistive Devices and Services for the Disabled: Auditory Signage and the Accessible City for Blind or Vision-Impaired Travelers. Signs (1998)
4. Loomis, J.M., Golledge, R.G., Klatzky, R.L.: Navigation System for the Blind: Auditory Display Modes and Guidance. Presence: Teleoperators and Virtual Environments 7, 193–203 (1998)
5. Heuten, W., Henze, N., Boll, S., Pielot, M.: Tactile wayfinder: a non-visual support system for wayfinding. In: Proceedings of the 5th Nordic Conference on Human-computer Interaction: Building Bridges, pp. 172–181. ACM, New York (2008)
6. Katz, B., Truillet, P., Thorpe, S., Jouffrais, C.: NAVIG: Navigation Assisted by Artificial Vision and GNSS. In: Workshop on Multimodal Location Based Techniques for Extreme Navigation, Finland, pp. 1–4 (2010)
7. Brilhault, A., Kammoun, S., Gutierrez, O., Truillet, P., Jouffrais, C.: Fusion of Artificial Vision and GPS to Improve Blind Pedestrian Positioning. In: Proceedings of the IFIP International Conference on Conference on New Technologies, Mobility and Security (NTMS), pp. 1–5 (2011)

Towards a Context Oriented Approach to Ethical Evaluation of Interactive Technologies

Sandra Burri Gram-Hansen, Henrik Schärfe, and Jens Vilhelm Dinesen

Department of Communication and Psychology, Aalborg University, Denmark
{burri,scharfe,jdinesen}@hum.aau.dk

Abstract. This paper explores and develops the notion of applying the ethical perspective of Danish philosopher and theologian K.E. Løgstrup, when designing and developing interactive technologies. The ethical reflections presented in this paper are currently considered in the development of Persuasive Learning Designs within the EU funded PLOT project, thus enabling this paper to support the argumentation with a practical example of integrating ethical considerations into the different stages of a design process.

Keywords: Ethics, Persuasive Design, Løgstrup, PLOT, e-Learning, Learning Objects.

1 Introduction

This paper explores and develops the notion of applying the ethical perspective of Danish philosopher and theologian K.E. Løgstrup, when designing and developing interactive technologies. The paper builds on previous research, in which the relevance of Løgstrup's ontological approach to ethics has been established[1], and provides a practical example of participatory design where this particular approach to ethics appears to offer valuable philosophical insight and considerations.

Previous research has raised the argument that the often applied utilitarian and deontological evaluation methods are insufficient when evaluating interactive computer technologies, as they fail to include reflections regarding the social reality of the users and the context in which the technology is applied and fail to acknowledge the mutual ethical responsibility between the designer and the user. This is problematic as designers are not necessarily able to foresee all possible use scenarios and should as such not be held solely responsible for unethical consequences of applying a technology[2, 3] In order to meet these shortcomings, Gram-Hansen suggests that the ontological perspectives and reflections presented by K.E. Løgstrup may be a beneficial supplement to ethical evaluations of interactive computer systems[1].

Løgstrup finds, that ethics, rather than being based on reason, is founded in what he calls *sovereign expressions of life*, which includes benevolence, open speech, trust, love and compassion – in other words human features that are generally considered ethically praiseworthy. He furthermore argues that we are born into ethics as a result of the dependency which exists between humans. As soon as humans interact, they influence each other's lives, and it is by interaction that ethics and ethical

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 628–631, 2011.

responsibility emerges. Furthermore, Løgstrup stresses that the perception of ethics is based on the contextual reality of the individual, i.e. ethics is considered an intuitive result of human nature, rather than moral rule based on reason, and the distinction between ethical and unethical actions are dependent on the specific situation and the social reality of the people involved in the interaction.

Considering Løgstrup's ethics when evaluating interactive technologies is not unproblematic. Firstly because Løgstrup argues that the perception of the ethical action is based on the intuition and social reality of the person performing an action – making it impossible for others to evaluate the action. And secondly because the notions of ethics which are presented by Løgstrup, originate from reflections concerning humans who are physically located at the same place thus sharing a common understanding for the characteristics which define the specific context. The ability to interact through technology has ended the necessity of interacting parties actually meeting, and when the interaction takes place between the designer and the user, they will most often not be found at the same place. As a result, the ethical perspective presented by Løgstrup cannot stand alone in the evaluation of interactive technologies, but should be applied as a third perspective in collaboration with utilitarianism and deontology. Løgstrup's contribution serves as a context oriented perspective, providing reflections and discussions which are as essential in interaction through technologies as they are for physical interaction between humans. In particular, Løgstrup offers valuable viewpoints concerning key concepts such as trust, credibility and interaction.

Whilst acknowledging that Løgstrup's ethical reflections are not exclusively applicable as an evaluation method, this paper will argue that the inclusion of key concepts and notions presented by Løgstrup may be highly beneficial to the design process - particularly if the process includes elements of value sensitive design (VSD) and participatory design. In order to exemplify how Løgstrup's ethical perspectives may be included in a design process, a selection of key concepts and core notions are considered in practice in the development of Persuasive Learning Designs within the EU project PLOT.

2 Putting Theory to Practice in the EU Project PLOT

The EU funded PLOT project (Persuasive Learning Objects and Technologies) was initiated in November 2010 and aims to develop a pedagogical framework for active engagement, based on persuasive design, as well as to demonstrate its value by creating tools and exemplars of adaptable, reusable learning resources. This pedagogical framework will incorporate persuasive design principles, which determine how to design ICT that can change people's behaviour and/or attitudes. The Persuasive Learning Designs on one hand incorporate known and well-tested didactic principles, and on the other hand offer implementation particularly designed to embody persuasive principles such as guided choices, simulations, competitions, and possibilities to share the experience with peers. The impact will be to generate more effective active e-learning resources and will give teachers the necessary tools to both create new and adapt existing resources to suit their needs. This will improve the quality of e-learning resources available for both teachers and students across[4].

The PLOT partner group consists of designers, developers, experts in learning and in persuasive design, as well as representatives from the four project work cases. As such, the partner group consists of both technical experts and members with little or no technical proficiency. Within the development process, this causes some classical difficulties in relation to common language and common understanding of the work process. Amongst the challenges to be met within PLOT, is the development of Persuasive Learning Designs for each of the four work cases. Each work case represents different learning material and different challenges, and do as a result require individual attention and consideration if the learning designs are to comply with not only learning theory but also with the notion of persuasive technologies and persuasive design.

In order to acquire the necessary contextual information about the four cases, the PLOT partners were asked to participate in an Inspiration Card Workshop which was held during the Euro PLOT consortium meeting on May 10^{th} and 11^{th} 2011. The aim of the workshop was to create a social context in which the individual case representatives were given the opportunity to explain and elaborate upon their individual challenges in teaching and learning, and for the additional members of the consortium to ask questions and reflect upon the different case scenarios.

The workshop setup was inspired by previous work of Halskov and Dalsgaard [5] and Davis [6], but was executed in a narrower and more focused version by applying technology cards. These cards were primarily examples from the Design With Intent Toolkit, which was developed by Dan Lockton, which focuses entirely on intentional design in technologies. [7]

In practice, the members of the consortium were divided into two groups, which each had case representatives from two cases as well as partners representing design and development teams. Supported by the inspiration cards, each of the cases were explained and discussed. In order to ensure direction within the discussions, the groups where required to complete the workshop by creating a poster for each case which described one of the challenges they were facing, and a suggestion to its solution.

3 Workshop Results

The notion of the workshop was to facilitate mutual understanding between the individual partners, which was not only vital to the case oriented development of Persuasive Learning Designs, but could also motivate a reciprocal responsibility between case representatives and project developers, that may be considered the very foundation of future ethical interactions.

During the workshop it became apparent that the approach did impose a greater mutual understanding between the members of the consortium, and that the inspiration cards served as inspiration during discussions as well as a mean to ensure that the topics discussed were the ones relevant to the development of persuasive learning designs.

More importantly however, the workshop constituted a contextual reality in which we as designers where able to include several of Løgstrup's key concepts in our development process. The workshop initiated interaction between developers and designers in a way which influenced the natural power balance between the participants and created a context in which terminology and professional background was less important compared to the ability to actively engage in discussion and idea development.

As mentioned, this particular approach is easily linked to the notions of VSD and participatory design, which are commonly credited as ways of ensuring the ethicality of the product being developed [8]. However, by adding Løgstrup's approach to ethics to this established methodological perspective, designers and developers who initiate and engage in participatory design processes, are provided with a philosophical foundation for the ethical evaluation, as well as a number of key concepts to consider as the participatory activities are defined. Whether development participation consists of workshops, interviews or agile development methods, interaction, reciprocal responsibility and the perception of power balance are of vital importance, and may in all cases be supported and nuanced by considering the philosophical works of K.E. Løgstrup.

References

1. Gram-Hansen, S.B.: Towards an Approach to Ethics and HCI Development Based on Løgstrup's Ideas. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5726, pp. 200–203. Springer, Heidelberg (2009)
2. Albrechtslund, A.: Ethics and Technology Design. Ethics and Information Technology 9(1), 63–72 (2007)
3. Gram-Hansen, S.B.: Persuasive Everyware - Persuasive Design Funderet i Kairos. Communication and Psycology, 88 (2010)
4. Schärfe, H.: EuroPLOT (2010), `http://c.aau.dk/plot.html` (cited)
5. Halskov, K., Dalsgaard, P.: Inspiration card workshops. In: Proceedings of the 6th Conference on Designing Interactive Systems. ACM, University Park (2006)
6. Davis, J.: Generating Directions for Persuasive Technology Design with the Inspiration Card Workshop. In: Ploug, T., Hasle, P., Oinas-Kukkonen, H. (eds.) PERSUASIVE 2010. LNCS, vol. 6137, pp. 262–273. Springer, Heidelberg (2010)
7. Lockton, D., Harrison, D., Stanton, N.: Persuasive Technology in a Wider Context. In: Oinas-Kukkonen, H., Hasle, P., Harjumaa, M., Segerståhl, K., Øhrstrøm, P. (eds.) PERSUASIVE 2008. LNCS, vol. 5033. Springer, Heidelberg (2008)
8. Davis, J.: Design methods for ethical persuasive computing. In: Proceedings of the 4th International Conference on Persuasive Technology. ACM, Claremont (2009)

Towards a Framework of Co-Design Sessions with Children

Emanuela Mazzone[1], Janet C. Read[1], and Russell Beale[2]

[1] Child Computer Interaction Group, University of Central Lancashire,
Preston PR1 2HE, UK
{EMazzone,JCRead}@uclan.ac.uk
[2] Advanced Interaction Group, School of Computer Science, University of Birmingham,
Birmingham, UK, B15 2TT
R.Beale@cs.bham.ac.uk

Abstract. In this poster we present a framework of the elements of co-design sessions with children. The involvement of children in the design process is important in order to understand their needs but it is often considered a complex practice. Considering a thorough appreciation of this practice as the basis for its accurate application, we addressed its complexity in a framework. To do so, we identified and organised elements that have an impact on co-design sessions in who, where, when, what and how dimensions. This theoretical framework aims to support novice practitioners in their decisions when coordinating co-design sessions.

Keywords: Interaction Design, Children, Co-Design, Framework.

1 Introduction

A User-Centred approach is recommended in the design of novel technology for children in order to reduce the discrepancy between the system conceptual model, defined by adult designers, and the mental model of children users. Designers usually refer to cognitive development stages of children to identify the users' capabilities and skills that need to be mirrored in the product requirements, i.e. [1, 2]. In addition to the reference to developmental models, it has become a common practice to involve children during the design and evaluation of products to make sure that the requirements are actually met. Engaging children throughout the design process can take several forms depending on the degree of participation children are allowed, from simply users to equal partners [3]. Over the last decade, there has been an increasing tendency to explore children's involvement in the early stages of the design process, i.e [5, 6, 4, 7]. The broad term 'co-design' means that users are asked to do things, instead of only being asked about a topic or observed in specific situations; this takes the designers to a deeper level of understanding of the user's way of thinking about the world [8]. When it comes to children, engaging them in design activities also helps overcome the limitations that arise from their not fully developed language and social skills, while at the same time encouraging their potential for creativity and imagination.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 632–635, 2011.

As concluded by Nesset and Large [9], when reviewing design projects that involved children, the advantages in terms of innovation and appropriateness of design outweigh negative factors, like cost in terms of time and resources needed. Therefore, ''*the real issue would seem to be […] how to more effectively engage children in the design process*'' [10]. In a similar way, in a special issue of the coDesign journal, all the studies agreed on the value of including children in the early stages of design and in the decision making process, but highlighted practical difficulties and the need for guides and directions for designers and researchers who want to organise and conduct meaningful participations [11]. More recently, [12] proposed a framework for supporting the choice of certain methods by relating the characteristics of children with the characteristics of design methods. As suggested by [13], research that aims to support the Interaction Design practice has to be based on a deep understanding of design as a complex human activity of inquiry and action. Taking all these research directions into account, we wanted to increase the awareness of all the elements involved in a co-design session that influence its output by organising them into a framework. The framework we present in this poster is of 'explanatory' [14] as it presents concepts or dimensions to consider in designing experiences, rather than presenting a set of steps or principles to follow.

2 Defining a Framework

From the analysis of the research reported in the literature and from our own experience in the field [15-17] we identified elements that are determinant in co-design sessions with children and need to be considered when managing resources for gaining access to children, for running the co-design sessions or for analysing the results. We consider these elements to also play a significant role in what we identified as critical aspects for the researchers, i.e.: determining researchers' expectations of the outputs, considering children's perspectives, pondering practical constraints, dealing with not-programmed incidents, and managing sessions' outputs. Our aim is to organise the complexity of co-design with children in a way that highlights the importance of each element in the handling of a co-design session. We understand the coordination of a design session as defining the design objectives and variables to involve in the session. Once researchers identify the objectives and the available resources as initial requirements, the decisions concern all those elements involved in the experience of the session. We adopted a *who-what-when-where-how* (WH/HW) structure as an easy-to-remember way to group these factors, also used by [18] for planning evaluation sessions. The 5 WH/HW dimensions are presented as tables to summarise the implications behind the decisions on the single variables involved in the sessions[1].

The *who* section includes consulting experts to support and validate the definition of activities. These experts can be teachers or developmental and educational experts whose advice is strategic for understanding children's skills and potential contribution. In the same section facilitators, like observers or assistants, are

[1] See http://193.61.245.74/interact-poster.pdf for a version of the framework described here.

considered – these need to share the same approach on the activities, understand their role, comprehend the activity goal, and the expected output. In this respect researchers should be aware of observers' effect and countercheck possible influences this may have on children's contributions. 'Children' do not appear as a specific variable of this section: since, as they are the main focus of the design session, their perspectives are addressed in other variables (i.e. experts advice, grouping, fun, distractions, etc.). For example, grouping strategies have an effect on the session's output: working in group can provoke discussion and chaining of ideas but may inhibit more introvert children, while individual work has the side effect of taking more time or leading to less ideas [12, 16].

The *where* section presents hints for space location, either familiar contexts for children but less controllable for researchers, or lab-type spaces that are less natural contexts for children. Concerning space distribution, children doing activities simultaneously in the same room can save time and resources but there will be an increase of chaos and distractions.

For the *when* section, options are presented for time management in terms of duration of activities and tasks, inclusion of breaks, and flexibility to adapt to unexpected events.

The *what* section includes a table with an alphabetic list of the most common techniques used in co-design sessions for children. Each technique has a brief description of its intended aim, some examples of possible variations, identification of the design stage of application, indication of the required children's skills, and a summary of the pros and cons of its employment. The introduction of props and of variants of creative techniques is important to trigger children's ideas and adapt to children's different ways of expressing themselves.

The *how* section considers ethics on privacy and security issues, together with tips on the different ways to brief the activity and hints on data collection. This last point is directly related to the analysis of outputs, which is essential to determine the session's success. For example, recording the progress of children's ideas and their explanations of the outputs can facilitate researchers' afterwards interpretation.

The aim of this framework is to support novice practitioners in reflecting on the implications of each element when taking decisions and outlining a detailed plan of their session. For example, if researchers want to involve children in the design of a computer game on a specific topic, they would check with the teacher whether that topic is suitable for the selected group of children to work on, or alternatively, look for a different group of children or adjust the topic to the available group. In a different example, if researchers are looking for a wide range of inputs in a limited time, they can ponder to organise a session with a whole group of children and recruit several assistant researchers to record the outputs rather than having small groups of children at a time to explore a topic in depth.

3 Conclusions and Future Work

The nature of the co-design with children practice and the endless variety of design situations and approaches makes it neither feasible nor sensible to create an exhaustive checklist for supporting novice practitioners in co-design sessions with children. In line with [13] and [19] views on research on design practice, design practitioners need concepts to 'think about' rather than 'guidelines' to tick off in a prescriptive way. In this paper we have organised elements involved in co-design sessions with children according to a WH/HW structure and presented their implications for the co-design session. This framework is the result of research based on literature and direct experience in the field and it provides theoretical references to use in practice. Once it is available to other practitioners of our research group, we will gather qualitative data to evaluate its effectiveness and refine its structure. Future work may also involve the creation of a practical tool based on this framework.

References

1. Antle, A.N.: The CTI framework. In: Proc. of TEI 2007, pp. 195–202. ACM, Louisiana (2007)
2. Gelderblom, H.: Designing software for young children. In: Proc. of IDC 2004, Maryland, pp. 121–122 (2004)
3. Druin, A.: The role of children in the design of new technology. BIT 21, 1–25 (2002)
4. Williamson, B.: The participation of children in the design of new technology. FutureLab (2003)
5. Bruckman, A., Bandlow, A.: Human-computer interaction for kids. In: The Human-computer Interaction Handbook, pp. 428–440. L. Erlbaum Associates Inc., Mahwah (2003)
6. Jones, C., McIver, L., Gibson, L., Gregor, P.: Experiences obtained from designing with children. In: Proc. of IDC 2003, pp. 69–74. ACM, Preston (2003)
7. Sluis-Thiescheffer, W., Bekker, T., Eggen, B.: Comparing early design methods for children. In: Proceedings of IDC 2007, pp. 17–24. ACM, Aalborg (2007)
8. Sanders, E.: Postdesign and Participatory Culture. Presented at the Useful and Critical: The Position of Research in Design, Finland (1999)
9. Nesset, V., Large, A.: Children in the information technology design process: A review of theories and their applications. L. & IS Research 26, 140–161 (2004)
10. Scaife, M., Rogers, Y., Aldrich, F., Davies, M.: Designing for or designing with? Informant design for interactive learning environments. In: Proceedings of CHI 1997, pp. 343–350. ACM, Atlanta (1997)
11. Woodcock, A.: Editorial. CoDesign 4, 193 (2008)
12. Sluis-Thiescheffer, R., Bekker, M., Eggen, J., Vermeeren, A., de Ridder, H.: Development and application of a framework for comparing early design methods for young children. Interacting with Computers 23, 70–84 (2011)
13. Stolterman, E.: The Nature of Design Practice and Implications for Interaction Design Research. International Journal of Design 2, 55–65 (2008)
14. Rogers, Y., Muller, H.: A framework for designing sensor-based interactions to promote exploration and reflection in play. IJ. of Human-Computer Studies 64, 1–14 (2006)
15. Kelly, S.R., Mazzone, E., Horton, M., Read, J.C.: Bluebells: a design method for child-centred product development. In: Proc. of NordiCHI 2006, pp. 361–368. ACM, Norway (2006)
16. Mazzone, E., Read, J., Beale, R.: Understanding children's contributions during informant design. In: Proc. of British HCI 2008, pp. 61–64. BCS, Liverpool (2008)
17. Mazzone, E., Tikkanen, R., Read, J.C., Iivari, N., Beale, R.: Integrating Children's Contributions in the Interaction Design Process. IJART (2012)
18. Markopoulos, P., Read, J., MacFarlane, S., Hoysniemi, J.: Evaluating Children's Interactive Products. Morgan Kaufmann, San Francisco (2008)
19. Schön, D.: The Reflective Practioner. Perseus (1983)

Towards a Semantic Modelling Framework in Support of Multimodal User Interface Design

Elena Tsiporkova, Tom Tourwé, and Nicolás González-Deleito

Sirris – Software Engineering & ICT Group,
A. Reyerslaan 80 – 1030 Brussels – Belgium
{elena.tsiporkova,tom.tourwe,nicolas.gonzalez}@sirris.be

Abstract. In this paper, we propose a semantic modelling framework to capture the available domain knowledge in the field of multimodal interface design and to support designers in their daily design tasks.

1 Introduction

Parallel with the evolution towards multimodal applications, the research toward the establishment of formal principles and guidelines for multimodal interaction design is gaining increasing interest and importance in recent years (e.g. [1,3,4]). However, as observed by Sarter [5], the existing guidelines mostly focus on high-level design objectives and do not provide support on how to map them to the needs of an actual application. They do not capture and reflect on the considerable practical experience and valuable expert knowledge that interface designers rely on during their daily activities. Moreover, a considerable gap exists between the theory (formal guidelines) and the practice of multimodal human interface design, as different experts might approach the same interface design tasks in different ways based on personal expertise, background and intuition.

Our aim in this article is to work toward bridging this gap via the application of semantic technologies (e.g. ontologies) for capturing the available domain knowledge in the field of multimodal interface design. There are several advantages associated with such an approach: it guarantees a uniform approach across different designers within the same organisation, allows for semantic inter-usability of the formal guidelines across different applications and domains, and is open to allow for knowledge evolution and growth.

2 Multimodal User Interface Design: Current Practices and Challenges

A recent survey [1] on the current state-of-the-art of multimodal interfaces covers exhaustively the foundations and features of multimodal interaction, current developments in modelling languages and programming frameworks, and existing principles and guidelines for multimodal interface design. Several authors worked on establishing formal principles for multimodal user interface design. Sarter [5] reviewed the existing design guidelines for multimodal information presentation,

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 636–639, 2011.

approaching the problem from the point of view of main decisions and considerations involved in multimodal interface design. She identified four themes of guidelines: 1) the *selection* of modalities; 2) *mapping* modalities to tasks and types of information; 3) the *combination*, synchronisation and integration of modalities; 4) *adaptation* of multimodal information input and presentation to accommodate changing task contexts and circumstances.

Sarter argued that most shortcomings in the aforementioned guidelines are due to the fact that there remain a substantial number of open research problems in the area of multimodal interaction design. However, we consider that guidelines resulting from research would always remain of a rather conceptual and less empirical nature and thus of little *practical* use. In our opinion, an extensive intuitive and empirical knowledge base exists already within the designer community, and the challenge is to develop methods, formal languages and frameworks that allow for capturing and exploiting this knowledge. The semantic modelling framework proposed in this work is an initial attempt in this direction. The selected approach is based on Sarter's insight of considering the interface design process from the perspective of the decision dilemmas designers face daily when executing their design tasks.

3 Semantic Modelling Framework

Our semantic modelling framework allows us to capture general *domain knowledge* and *expert knowledge*, the latter is referring to Human-Computer Interaction (HCI) knowledge related to multimodal application design. Both domain and expert knowledge are described via an ontology, a formal representation of knowledge by a set of key domain concepts and the relationships between those concepts.

3.1 Competency Questions and Domain Concepts

To determine the scope of an ontology and identify the questions that an ontology-based knowledge repository should be able to answer, we need to state so-called *competency questions* [2]. Subsequently, from these questions the information that needs to be contained within the ontology, i.e. the concepts and the relationships between concepts, can be derived.

We consider that at this stage of the research, our ontology should be able to answer the following competency questions, which are inspired by the major themes identified by Sarter [5] (see above): 1) Which input and output modalities are available to the user of an application? 2) What are the different factors that affect the use of particular input and output modalities? 3) What are the appropriate (combinations of) modalities to support users in a particular task?

Based on this list of questions, the ontology will include information on users, their context and tasks, the devices they use and their specific capabilities, and usage characteristics and constraints of particular user-interface modalities.

Fig. 1 shows the high-level *domain concepts* of our ontology. For instance, the class **User** represents a user of an application, which itself is represented by the class **Application**. A class **Location** represents the physical location where the user is performing his activities and using the application. Locations are considered as being an **IndoorLocation** or an **OutdoorLocation**. The class **Activity** represents the activities that a user can engage in, subdivided into **PrimaryActivity** and **SecondaryActivity**.

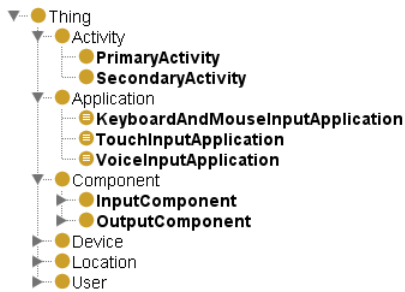


Fig. 1. High-level domain concepts in the semantic modelling framework

These concepts are related through relationships. Some relationships are defined through *property chains* which enable them to be automatically inferred from other relationships. For example, if we know that application *A* is used by user *U*, and that this user is located in location *L*, we can infer that application *A* is used in location *L*. More formally, $A\ used\_by\ U \wedge U\ located\_in\ L \rightarrow A\ used\_in\ L$.

3.2 Modelling Domain Knowledge and Design Guidelines

We model *domain knowledge* by defining subclasses for the key domain concepts, specifying necessary and sufficient conditions as appropriate. We consider domain knowledge to be any factual information about users, applications and devices that potentially influences the decision about which modality to provide. This includes obvious information such as the specific input/output capabilities of a device, but also information such as physical and social aspects of the user's working environment, or particular aspects of the nature of the activity (e.g. primary and secondary tasks). For example, we consider that **Microphone** is a subclass of **VoiceInputComponent** which in itself is a subclass of **InputComponent**, and that **PrivateWorkingSpace** is an **IndoorLocation** that is quiet (by adding a condition **has_noise_level some Quiet**).

Design guidelines capture the expertise and experience of the HCI practitioners. They describe applicability conditions and constraints for the use of a particular multimodal interface. For instance, interaction through a vocal command can be suggested if the following conditions and constraints are met: 1) the application runs on a device that has access to a microphone; 2) the application is used in an environment with limited noise, since the accuracy of voice technology is heavily dependent on environmental noise conditions (e.g. background noise); 3) the application is used in a private environment, since controlling an application by voice might disturb surrounding colleagues; 4) the user only uses the application as a support for his primary activities which require the use of both hands, i.e. the application itself is not the primary focus of the user.

We model this by introducing a subclass **VoiceInputApplication** of **Application**, representing applications that could use vocal commands, and by defining a necessary and sufficient condition representing the above conditions as follows:

(has_access_to some VoiceInputComponent) and **(used_in some PrivateWorkingSpace)** and **(used_while_performing some (PrimaryActivity that requires some BothHands))**

4 Conclusions

This paper presents an initial attempt to formally model and exploit relevant domain knowledge and expertise in support of selecting appropriate modalities during the human machine interface design process. The work presented here is inspired and results from our interactions with HCI practitioners. Besides further refinement of the presented semantic modelling framework, future research includes considering other competency questions that our ontology could support, expanding toward broad range of working contexts and types of users, and supporting synchronisation of modalities in time. All this research will be performed in close collaboration with HCI practitioners.

Acknowledgements. This research is supported by the ARTEMIS JU and IWT (www.iwt.be) through the SMARCOS project (www.smarcos-project.eu), and by Innoviris (www.innoviris.be) through the VariBru project (www.varibru.be).

References

1. Dumas, B., Lalanne, D., Oviatt, S.: Multimodal interfaces: A survey of principles, models and frameworks. Human Machine Interaction, 3–26 (2009)
2. Gruninger, M., Fox, M., et al.: Methodology for the Design and Evaluation of Ontologies. In: Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing, IJCAI, vol. 95 (1995)
3. Oviatt, S.: Multimodal interfaces. In: The Human-computer Interaction Handbook: Fundamentals, Evolving Technologies and Emerging Applications, pp. 286–304 (2003)
4. Reeves, L., Lai, J., Larson, J., Oviatt, S., Balaji, T., Buisine, S., Collings, P., Cohen, P., Kraal, B., Martin, J., et al.: Guidelines for multimodal user interface design. Communications of the ACM 47(1), 57–59 (2004)
5. Sarter, N.: Multimodal information presentation: Design guidance and research challenges. International Journal of Industrial Ergonomics 36(5), 439–445 (2006)

Towards an Experimental Framework for Measuring Usability of Model-Driven Tools*

Jose Ignacio Panach[1], Nelly Condori-Fernández[2], Arthur Baars[1], Tanja Vos[1], Ignacio Romeu[1], and Óscar Pastor[1]

[1] Centro de Investigación en Métodos de Producción de Software ProS, Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain
{jpanach,abaars,tvos,iromeu,opastor}@pros.upv.es
[2] Faculty of Electrical Engineering, Mathematics and Computer Science, University of Twente, Information Systems Group, 7500 AE Enschede, P.O. Box 217, The Netherlands
nelly@pros.upv.es

Abstract. According to the Model-Driven Development (MDD) paradigm, analysts can substantially improve the software development process concentrating their efforts on a conceptual model, which can be transformed into code by means of transformation rules applied by a model compiler. However, MDD tools are not widely used in industry. One of the reasons for this poor adoption is the lack of usability of MDD tools. This paper presents a framework to evaluate the usability of such tools. The framework will be used as a basis for a family of experiments to get clear insights into the barriers to usability that prevent MDD tools from being widely adopted in industry.

Keywords: Usability, model-driven development, evaluation framework.

1 Introduction

MDD tools can offer many benefits for a software developing company: reduced costs, reduced development time, higher quality and higher customer satisfaction [7]. However, the wide acceptance of MDD is not yet a reality. One reason for this situation is the lack of usability of MDD tools. The main contribution of this paper is an empirical framework to perform a set of usability evaluations for MDD tools. By experimental framework we mean the definition of a process and the elements involved in it with the aim of performing evaluations unambiguously. The framework will be used to perform the evaluation of several MDD tools knowing clearly the required elements and the stages to perform the experiment.

Several authors such as Fiora [2] and Kostiainen [5] have proposed experimental frameworks for measuring system usability. If we focus our research on MDD tools,

* This work has been developed with the support of MICINN, GVA and ITEA 2 call under the projects PROS-Req (TIN2010-19130-C02-02), ORCA (PROMETEO/2009/015) and UsiXML (20080026).

little work has been published about their usability. We can only find usability frameworks defined for evaluating usability in Computer Aided Software Engineering (CASE) tools in general, such as the work of Senapathi [8]. However, frameworks defined for CASE tools are not always valid for MDD tools. In contrast to conventional CASE tools that are oriented to software development based on design and programming, MDD tools have to cope with specific features where the modeling and the programming perspective become intertwined. MDD tools have as peculiarity that they have an underlying software delevopment method which must be easily understandable to the user. Studying related work, we conclude that more work must be done on the usability evaluation of MDD tools to become fully accepted in industry. The paper is divided into these sections: 2nd describes the framework, 3th applies the framework to a specific MDD tools, and 4th presents the conclusions.

2 Empirical Framework to Evaluate the Usability of MDD Tools

The empirical framework is composed of a usability evaluation model and an experimental process, which have been designed using Wohlin's proposal [9].

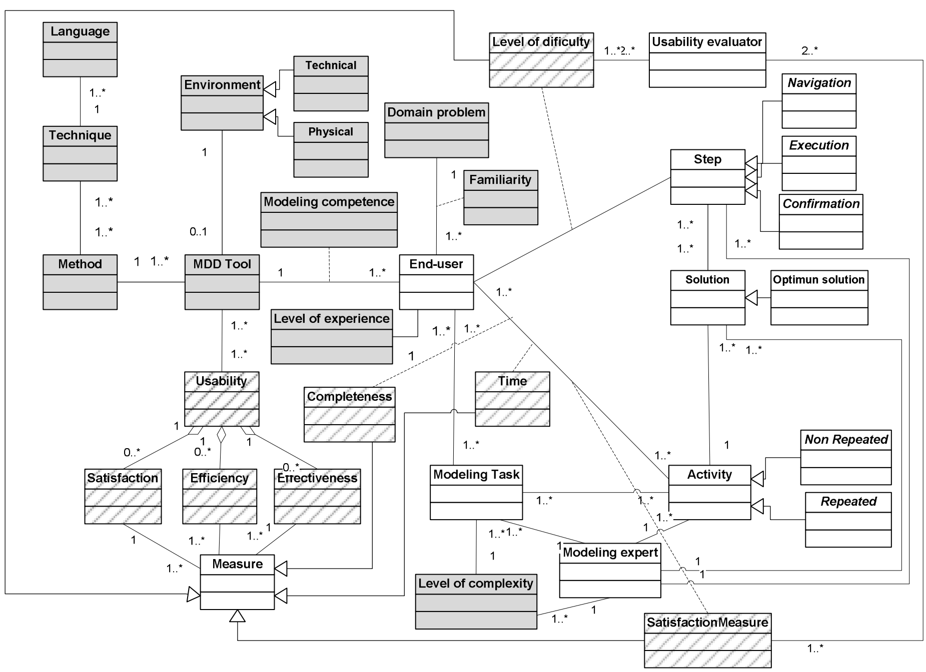


Fig. 1. Usability Evaluation Model

The **usability evaluation model** identifies the most relevant elements for evaluating the usability of MDD tools (Figure 1). According to Figure 1, we aim to evaluate the usability of MDD tools by means of: satisfaction, efficiency and effectiveness, such as ISO 9241-11 proposes [4]. In summary, this model represents:

the method that underpins the MDD tool, the end-user features, and the activities that the end-user will execute in the experiment. Classes with a grey background represent elements that can be changed and controlled in the experiment and classes with a background of diagonal lines represent elements that can be quantified with one or more measures.

The **experimental process** focuses on the stages that compose the process to perform the usability evaluation according to the usability model. There are four stages [9] (Figure 2): (1) **Definition**: This determines the foundation of the experiment; (2) **Planning**: This specifies how the experiment is conducted. (3) **Operation:** In this step, the researcher performs the experiment and collects the needed data; (4) **Analysis**: In this step, the researcher interprets the experiment data. Each one of these stages is divided into substages, such as Figure 2 shows.

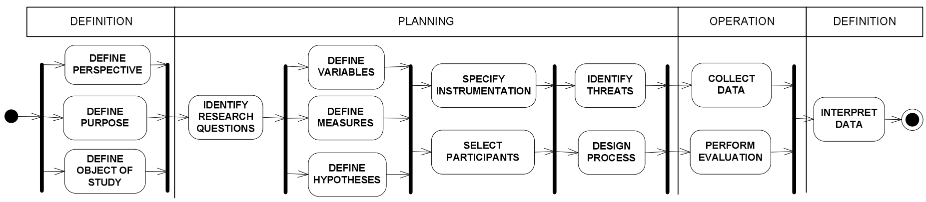


Fig. 2. Process to evaluate usability

3 Applying the Usability Framework to a Specific MDD Tool

This section explains the usability evaluation that we performed with the proposed framework applied to an MDD tool called OLIVANOVA [1]. We focus our study on effectiveness and efficiency, relegating the satisfaction to a future experiment. Next, for space reason, we detail only some elements that compose the Planning stage (Figure 2): hypotheses, variables and participants. However, it is important to note that the Usability Evaluation Model (Figure 1) is instantiated throughout all the substages that compose the Definition and Planning stages (Figure 2).

We have one null hypothesis for efficiency and another for effectiveness:

- **H_{10}:** When using OLIVANOVA for modeling tasks with different levels of complexity, the efficiency is the same independently of users' experience.
- **H_{20}:** When using OLIVANOVA for modeling tasks with different levels of complexity, the effectiveness is the same independently of users' experience.

We have identified dependent **variables** and independent variables [9]:

- **Dependent variables:** Efficiency and effectiveness. **Efficiency** was measured by task completion percentage in relation to the time spent to perform a task. **Effectiveness** is the level of completeness reached in every task. This variable was calculated by two measures: 1) the percentage of tasks carried out correctly; 2) the percentage of correctly performed activities that were carried out optimally (without any difficulty compared to experts).
- **Independent variables:** The level of complexity of the tasks, the modeling competence, and the level of experience using MDD tools.

We used three groups of participants:

- **Type I (Experts)**: Experienced using the evaluated tool. This group was composed of researchers of the ProS center of the Technical University of Valencia.
- **Type II (Medium)**: Experienced using similar tools. This group was recruited from the regional Valencian public administration who are familiar with open source MDD tools like Moskitt [6].
- **Type III (Novice)**: No experience with the tool nor with similar tools. This group was composed of engineers from the Technological Institute of Computer Science who are familiar with Object-Oriented concepts but not with any modeling tool.

We used four users from each group to work with a balanced number of users. Instruments used to perform the experiment can be found at [3]. Once the experiment was finished, we analyzed the results of the efficiency and the effectiveness. **Efficiency** was measured by task completion percentage in relation to the time spent doing a task. According to an ANOVA test, we reject the null hypothesis. However, if this analysis is carried out excluding the group of experts, there are no differences in the mean efficiency scores with the other two types of users. **Effectiveness** was measured in terms of modeling task completion percentage and percentage of correct tasks that were carried out **optimally**. With the ANOVA test, we found significant differences for both the tasks of medium level of difficulty and high level of difficulty in the three groups of users. Therefore, we conclude that H_{20} is not satisfied.

4 Conclusions and Future Work

This paper proposes an empirical framework to evaluate the usability of MDD tools. The framework aims to replicate the usability evaluation of MDD tools in similar conditions. With the purpose of evaluating our framework; it was applied to a MDD tool called OLIVANOVA. The results of the evaluation demonstrate that changes must be applied to OLIVANOVA to improve the users' effectiveness and efficiency. As future work, we want to repeat our usability test with more users and considering the user's satisfaction. Moreover, we want to assess the framework with other MDD tools different from OLIVANOVA.

References

1. CARE Technologies S.A, `http://www.care-t.com`
2. Fiora, T.W., Au, B.S., Warren, I., Dobbie, G.: Automated Usability Testing Framework. In: Proc. 9th Australasian User Interface Conference, AUIC 2008 (2008)
3. Instruments, `http://hci.dsic.upv.es/testOLIVANOVA/`
4. ISO 9241-11 Ergonomic requirements for office work with visual display terminals (1998)
5. Kostiainen, K., Uzun, E.: Framework for Comparative Usability Testing of Distributed Applications. In: Security User Studies: Methodologies and Best Practices Workshop (2007)
6. Moskitt, `http://www.moskitt.org/`
7. Selic, B.: The Pragmatics of Model-Driven Development. IEEE Software 20, 19–25 (2003)
8. Senapathi, M.: A Framework for the Evaluation of CASE Tool Learnability in Educational Environments. Journal of Information Technology Education 4, 61–84 (2005)
9. Wohlin, C., Runeson, P., Höst, M.: Experimentation in Software Engineering: An Introduction. Springer, Heidelberg (1999)

TROCAS: Communication Skills Development in Children with Autism Spectrum Disorders via ICT

Margarida Lucas da Silva[1], Carla Simões[2,3], Daniel Gonçalves[1,4], Tiago Guerreiro[1,4], Hugo Silva[1], and Fernanda Botelho[3]

[1] IST-UTL
Av. Rovisco Pais, 49, 1050-001 Lisboa, Portugal
margarida.lucas.silva@ist.utl.pt
[2] FCSH – UNL
Av. de Berna, 26-C, 1069-061 Lisboa, Portugal
m.carla.simoes@gmail.com
[3] ESE – IPS
Campus do IPS, Estefanilha, 2914-504 Setúbal, Portugal
fernanda.botelho@ese.ips.pt
[4] INESC-ID
R. Alves Redol, 9, 1000-029 Lisboa, Portugal
{tjvg,daniel.goncalves}@vimmi.inesc.pt

Abstract. Autism spectrum disorders (ASDs) are a triad of disturbances affecting the areas of communication, social interaction and behavior. In educational contexts, without appropriate intervention methodologies, these limitations can be deeply disabling. Our research promotes the communicative competence of children with ASDs It extends the current state-of-the-art in the field, both in terms of usability for the educators and functionality for the end-users. We performed a long-term study, and results suggest that the proposed approach is effective in promoting the development of new interaction patterns.

Keywords: Communicative competence, Autism spectrum disorder, Assistive technologies, Computer-mediated communication, Multimedia platform.

1 Introduction

Autism Spectrum Disorders (ASDs) are included in the global developmental disorders group, and considered to be severe and precoce neuro-development disorders that have no known cure and persist throughout life [1]. In this study, we have looked for evidence about the benefits of using a novel multimedia platform, designed specifically to develop the communicative competence in children with ASD. We have mainly focused on making it fully customizable, namely in being: (a) Appealing to the target group; (b) Easy to use both by the tutor and by the end-users; (c) Portable and easy to deploy; (d) Adaptable to the severity degree of the target users; (e) Configurable content-wise to the end-user's development stages; and (f) Evolutive according to their age and school grade.

The paper is organized as follows: Section 2 shows an overview of ASD, unveiling the need for tools to promote the communicative competences in the educational

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 644–647, 2011.

context, which led us to the development of a novel platform, presented in section 3, capable of providing the best experience for children and their tutors. In sections 4 and 5 we analyze the experimental results and outline the main conclusions.

2 Motivation

Communication competences and social interaction are one of the areas where individuals with ASD present a higher compromise. At the education context level, an ASD frame [2] requires a set of specialized intervention measures that can attenuate the communication and integration difficulties faced by these children. Their communication abilities can be improved through specialized intervention models [3].

Currently, there is a lack of assistive technologies for the specific development of the communicative competence in children with ASDs. Existing solutions have been developed in a more enabling and functional perspective, centered in the individual himself. Examples of possible applications are: the Zac Browser, an enabler for Internet access and multimedia activities; and the Grid 2, a tool for Human-Computer Interaction. Both are aimed at different target populations.

The integration and learning difficulties faced by these children in special education schools has led to new guidelines regarding the action models to employ [4]. Currently, these children are integrated in regular education schools where they learn social inclusion. Our work is focused on a real world problem and application, and on the improvement of current practices.

3 Proposed Approach

Based on the practical needs of children and their tutors, we have created a fully customizable platform, with the resources that are considered beneficial for the learning environments of children with ASD, and with potential to promote the acquisition of new competences at the communication level. Under tutor supervision, the children created the overall software appearance in a way that better relates their perceived visual information with the underlying context information. From the literature revision in the field, regarding communicative competences promotion in autistic children and adequate stimuli, the platform was designed to support: (a) photos; (b) audio; (c) videos; (d) message board; (e) online digital book library access; and (f) connection to a story retell tool, MS Photostory.

All sections allows the preference classification through the “Like” and “Dislike” buttons (**Fig. 1**). In a very simplified way, the child can express and share his/her opinion with others, which can in turn visualize the opinions and recognize the colleagues that have expressed them. Together with the message board (**Fig. 1**), these are core and distinctive features of our proposed approach to the problem. The message board was developed so that users can exchange messages among themselves, recurring to writing, to symbols, or a simultaneous combination of both. The entire platform is supported by Web standards, which makes it customizable and easily adaptable to other needs that children may have. Design considerations have taken into account two different user roles: (a) end-users that use the platform; and (b) tutors that manage the platform.

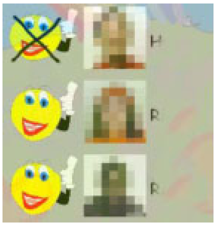


Fig. 1. Preference classification list

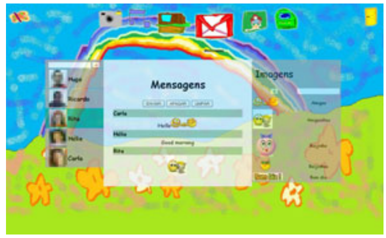
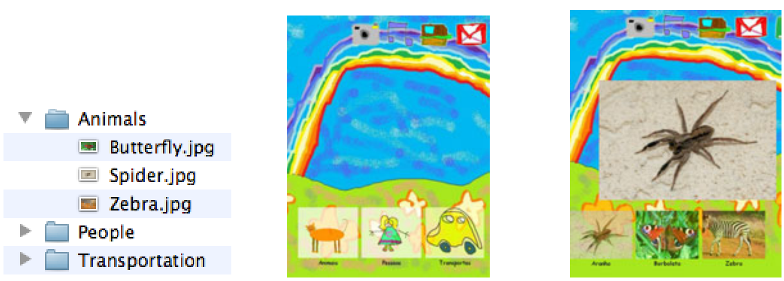


Fig. 2. Screenshot of the message board

A core aspect of the proposed platform is flexibility, that is, it can be easily adapted and customizable to the needs of each individual. Tutors can fully adjust the platform according to their goals, without the need of special knowledge regarding the use of the computer; contents are managed through the standard OS file management system, just like any other files. Accessing the TROCAS folder, the tutor can add, change, delete, and categorize the different contents through the folders. The platform automatically detects and organizes all content on the visual interfaces. Fig. 3 illustrates this process: in the folder where pictures are managed, the tutor created three categories: Animals, People, and Transportation (Fig. 3a.). The platform automatically detects these folders and organizes them on the visual interface as categories, using the name of the folder used as category descriptor (Fig. 3b.). The fact that the file names are used directly also allows the tutor to easily change the description of each content element, and facilitates the localization to other languages.



(a) Folder structure (b) Content organization in the platform

Fig. 3. Content management and categorization in the proposed platform

4 Experimental Results

Tests were conducted with 3 children, all with 12 years old; two of them integrated in a 5th grade class and a third in a 6th grade class, and all with individualized educational programs. Tutors were in charge of assessing profiles and progression according to his expert opinion. The implementation plan was: (a) a first stage of 2 weeks, which allowed the content fine tuning of the contents, platform operation, and the intervention plan; and (b) a second stage of 12 weeks, in which the established intervention plan was followed. Collected data revealed a consistent set of changes in the students communicative behaviours. The message board used as a way of communicating among the different agents in this process (counterparts and tutors) seems to favour the capability of understanding messages and orders. Students revealed a higher level of initiative and autonomy in exchanging messages, as well as a greater capacity of interpreting information and producing adequate responses.

It is important to enhance the fact that one of the students, in a final stage of the school year, by his/her own initiative brought a set of notes with the favorite TV characters, and asked to build a story using the resources made available by the platform. The communicative interactions between counterparts and tutors also became more frequent; the test subjects spoke frequently about the different contents of TROCAS, issuing their opinion about a given music, video, photo, etc. Of noticeable importance is the fact that one of the children, by listening to a colleague speak while viewing one of the videos expressed that: "*R. can speak after all*".

At an interpersonal relations level, the tool shows results in promotion of the relation and social interaction between counterparts. In general, students have always shown interest in using the platform, with clear increases in autonomy and proficiency not only using the platform but also using the computer, in general. Several comments and manifestations were registered, such as: (a) Requests to share TROCAS with other colleagues, tutors, and technicians; (b) Requests to use TROCAS event out of their assigned schedule; and (c) Refuse to attend classes to keep using TROCAS.

5 Conclusions

The use of information and communication technologies (ICT) is extremely important, as it can help reshape the teaching procedures, and make them more adequate to the reality of the subjects with special education needs. The inclusion of children with ASD in regular education schools is growing exponentially. This fact motivates the need to find new educational responses since, through adequate intervention models, the children can develop basic social competences, indispensable for an effective integration in the educational and social contexts. Experimental results collected during the period of deployment of TROCAS, show a clear relation between the intensive and guided use of the platform, and an improvement in the communicative competences among the test population. Future work will focus on implementing the tool in interfaces that may be more natural for the children.

References

1. APA: DSM.IV-TR. Quick Reference Guide for Diagnosis Criteria. Climepsi Editors (2002)
2. Jordan, R.: ASD in the Early Years: A Guide for Practitioners. QEd Pub. (2000)
3. Hewitt, S.: Specialist Support Approaches to ASD Students in mainstream settings (2005)
4. UNESCO ED-94/WS/18: The Salamanca Statement and Framework for Action on Special Needs Education. UNESCO (1994)

Usability Assessment of a Multimodal Visual-Haptic Framework for Chemistry Education

Sara Comai and Davide Mazza

Politecnico di Milano,
Department of Electronics and Information (DEI),
Piazza L. da Vinci, 32,
I-20133 Milan, Italy
{sara.comai,davide.mazza}@polimi.it

Abstract. In this work, we assess the usability of a virtual environment where the force of interaction between the electrostatic field around the molecule and a charge associated to the proxy of a haptic device can be felt. Feedbacks to user are provided in a multimodal visual and haptic way, and auxiliary information are also rendered.

Keywords: Haptics, chemistry education, usability evaluation.

1 Introduction

In this work, we assess the usability of a virtual environment described in [3, 4] enabling users to probe the electrostatic field around it with a charge associated with a haptic device. The force of interaction can be felt via the haptic device, and feedbacks are provided in a multimodal way, visual and haptic.

The framework is fully described in [3]. It provides the user with a virtual environment showing a 3D representation of the molecule; different modalities of visualization are possible among the typical chemical representation of a molecule three-dimensional structure, e.g. space-filling, ball-and-stick, van der Walls surfaces, etc. Beside the graphical representation, the framework allows to *feel* the electronic fields' interactions by means of a haptic device (in our case, the Sensable PHANTOM ® Omni [1]). The graphical environment shows the current position of the haptic proxy, as a yellow circle, to indicate which part of the space the user is interacting with. Auxiliary visual elements are shown to provide the user with a multimodal way of interaction, in order to integrate the information received in force, e.g., the plot of the electrostatic field along the direction connecting the proxy position and the center of the molecule, which can help in detecting maximum/minimum points of the electrostatic surface, or the force field on the molecule surface (represented by means of different colors). Figure 1 shows a screenshot of the GUI of the framework. The 3D molecule and the proxy pointer are clearly visible. The plot of the electrostatic field is shown in the upper-left corner of the picture. Further details on framework's visual and haptic capabilities can be found in [4].

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 648–651, 2011.

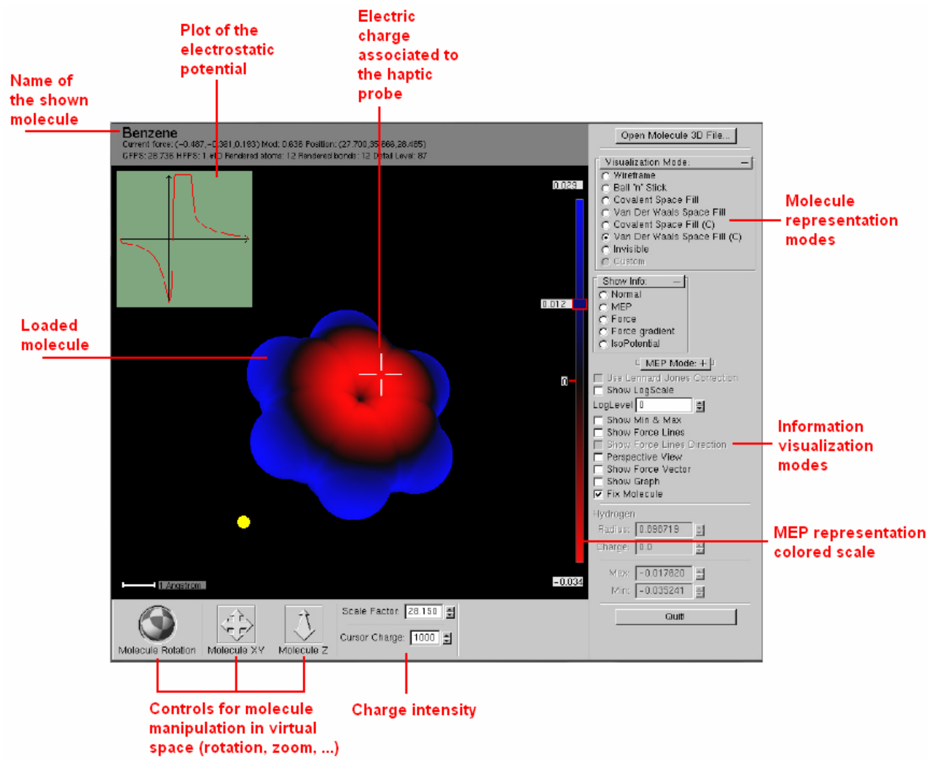


Fig. 1. A screenshot of the framework

2 Design of the Usability Test

Potential users of the application have been identified at first. Among the candidate users (i.e., students attending a university-level chemistry course, teachers, and chemistry researchers), being the framework mainly designed for *e-learning* activities, only students have been involved in this first evaluation study. The usability evaluation consisted of three stages:

- *pre-questionnaire administration*, sent to the users in order to collect the background information and knowledge about the topics. Beside providing some personal data on their studies and background, users were asked to rate their knowledge about both inter-molecular forces and haptic devices, on a scale ranging from 1 (*I do not know at all*) to 5 (*I know it very well*);
- *tasks' execution session*, where each user is asked to individually perform *three tasks*, consisting in identifying with the framework a set of specified electrostatic properties of a given molecule. While executing the tasks, the participants were asked to *"think-out-loud"*. They were required to fill in a questionnaire regarding the task they are performing (the adopted questionnaire form is ASQ [5]). The following task have been performed, by following a set of instructions provided on paper:

- Task 1 - *Critical points detection*: the charge is used to determine the maxima/minima of the potential in the space surrounding the molecule, giving a sensation of falling into an electrostatic surface "hole" when such points are found.
- Task 2 - *Detecting polar molecules*: different molecules exert a different influence on a positive or negative charge, and in different zones of the surrounding space;
- Task 3 - *Feeling the anisotropy of the interaction*: force effects by the molecule on the charge can change w.r.t. the charge-molecule distance and direction of approaching;

Execution of all the tasks has been supervised by us, users' behaviours observed and notes taken about what they said. The time taken by each user to complete each task has been also considered for the evaluation;

- *post-questionnaire administration*, providing a final evaluation of the application, by the usage of the System Usability Scale (SUS) [2]). Users were asked to answer a questionnaire to evaluate general and haptic-specific features of the application, and to finally provide an overall judgement of the experience.

A total of 5 users participated to the evaluation study of the developed chemical framework. All were students of the MSc in Computer Engineering at Politecnico di Milano. The group consisted of 2 females and 3 males with the average age of 21.6 years.

3 Results and Discussion

The pre-questionnaire results show that they have very little knowledge about haptic devices and an average knowledge about inter-molecular forces. With respect to chemistry softwares' background knowledge, only one participant was already using another chemistry application. The mean values of the times taken to complete each task, and the ASQ score, have been calculated. The ASQ score [5] includes the efficiency, effectiveness, satisfaction and use of haptic device usability aspects: the smaller the value, the better evaluated are these aspects by the participants. Users' feedbacks were collected regarding the different modalities of visualization and visual elements added to make haptic effects more intuitive. Based on these feedbacks, the most useful visual clues were determined as well as the problematic ones. Problems can be categorized as follows:

- *Semiotic*- problems related to the meanings of the messages of the application;
- *Cognitive* - cognitive effort of the user while using the application;
- *Graphics* - problems related to the graphical choices of the application;
- *Haptic* - problems related to the haptic features of the application including spatial problems and a haptic device.

Using the post-questionnaire, the score of the *haptic features* has been calculated using the same method as one for the SUS questionnaire [2]. The mean score of the haptic features of the application (43) is above average which shows that the participants found that the haptic features of the application are useful and allows user to perform the tasks in an effective and efficient way.

The SUS score of the *general features* of the framework is 81.5 which falls in the acceptability range of the usability of application and it is even higher than the mean score (76.2) of GUI interfaces [2] that were evaluated by SUS. The respondents rated the user-friendliness of the application in average *Good*, which corresponds also to the results of the SUS score.

Also the score of the *final feedbacks* is very high (4.3/5), confirming that the users have understood the purpose of the application, they found it attractive and would even recommend it to the colleagues.

During the test administration and as final feedbacks, students reported us the easiness by which they understood the explained concepts, thus leading to an increasing interest in the subject. Teachers also appreciated the improved awareness of the phenomena they usually explain only in theory.

4 Conclusions and Future Work

The framework has been tested by different users (students) on various molecules stored in a repository associated with the framework.

All the users recognized as the most important benefit the possibility of combining the typical visualization of chemical data with the rendering of the *feeling* of nanoscale/atomic interactions that can improve the understanding of real phenomena. As future work, given that students greatly appreciated it and they were literally enthusiastic when facing the novelty of such a tool, we plan to introduce it in teaching activities, i.e. in chemistry course as auxiliary tool to integrate classical lectures.

Acknowledgements. We would like to thank Margarita Lukjanska for her precious contribution on the usability assessment activity.

References

1. Sensable Phantom Omni. Web site, http://www.sensable.com/haptic-phantom-omni.htm
2. Bangor, J.T.M.A., Kortum, P.T.: Determining what individual SUS scores mean: Adding an adjective rating scale. Journ. of Usability Studies 4(3), 114–123 (2009)
3. Comai, S., Mazza, D.: A haptic-enhanced system for molecular sensing. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5727, pp. 493–496. Springer, Heidelberg (2009)
4. Comai, S., Mazza, D.: Haptic and visual rendering for multi-modal exploration of molecular information. In: Proceedings of BSC HCI (2010)
5. Lewis, J.R.: Psychometric evaluation of an after-scenario questionnaire for computer usability studies: The ASQ. SIGCHI Bulletin 23(1), 78–81 (1991)

Usability Planner: A Tool to Support the Process of Selecting Usability Methods

Xavier Ferre[1] and Nigel Bevan[2]

[1] Universidad Politecnica de Madrid, Campus de Montegancedo, 28660 - Boadilla del Monte (Madrid), Spain
xavier.ferre@upm.es
[2] Professional Usability Services, 12 King Edwards Gardens, London W3 9RG, UK
mail@nigelbevan.com

Abstract. There is increasing pressure on developers to produce usable systems, which requires the use of appropriate methods to support user centred design during development. There is currently no consistent advice on which methods are appropriate in which circumstances, so the selection of methods relies on individual experience and expertise. Considerable effort is required to collate information from various sources and to understand the applicability of each method in a particular situation. Usability Planner is a tool aimed to support the selection of the most appropriate methods depending on project and organizational constraints. Many of the rules employed are derived from ISO standards, complemented with rules from the authors' experience.

Keywords: UCD method selection, development process, usability integration in systems development, standards.

1 Introduction

UCD (User Centred Design) approaches and methods for systems development are used to help build systems that are easy to use. While there is general consensus on the need for a user-centred approach to design, there is a great diversity of potential methods. Existing literature on usability methods offers little guidance on which specific methods should be used in a given situation.

In order to adopt a UCD focus in development, usability has to be considered throughout the different development stages. The specific UCD methods to apply at each stage must be selected, depending on the relevant project and organizational constraints. This is a challenge because systems development teams and organizations can experience difficulty when trying to understand UCD methods and their role in the overall development process [6].

Existing advice on UCD method selection starts with the method, rather than the purpose for which the method is used. Senior UCD professionals get acquainted during the development of their career with the different subtleties of each method that make it more advisable in certain situations, but other groups (junior UCD consultants, HCI students or even software developers with some HCI training) face

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 652–655, 2011.

important difficulties for choosing the most appropriate methods for a particular project. They need to gather information present in different sources, and they have to deal with different terminology across authors in terms of process descriptions, and in terms of criteria for considering the suitability of each method.

The ISO PAS (Publicly Available Specification) 18152:2003 [4] provides an overall view of the life cycle for developing systems for human use, highlighting the base practices associated with the different processes involved. It does not provide specific guidance on method selection but it serves as an overall process framework where to fit advice on UCD methods integration in the systems development process.

The ISO TR (Technical Report) 16982 [5], on the contrary, specifically addresses the issue of UCD method selection. It does not work with individual methods to be applied in a specific situation, but with categories of them. It has an important shortcoming from a professional perspective, in being “somewhat academic” (according to the Special Interest Group in Software Testing from the British Computer Society [1]). Furthermore, the usage of this technical report is not straightforward, since it requires moving between different tables and gathering all the relevant information that is scattered in the document.

Usability Planner [7] is an online tool that supports the process of UCD method selection. It aims to ease the present difficulties found in considering the different knowledge sources, integrating the overall view in ISO PAS 18152:2003 with the recommendations present in ISO TR 16982, and supplementing them with the authors’ own experience.

2 Method Selection with Usability Planner

The difficulties traditionally found in UCD method selection arise from the difficulty of dealing with a diversity of issues that may influence the choice of methods. In a common scenario, UCD has already been identified in the organization as necessary for improving the quality of the system to produce, and there is a decision on the – typically limited – resources to dedicate to UCD in the project.

In these circumstances only a few UCD methods may be chosen, with the aim of maximizing their impact on the final usability level of the system produced. The sequence for making this choice using Usability Planner is:

1. Establish the overall background to consider: Either a UX (User eXperience) - UCD background, or a software engineering (developer) one. Depending on the choice a different set of stages into which to allocate the methods is offered.
2. Identify the specific project stages for which methods are to be selected. At this point the user may optionally prioritize the stages according to business benefits or risk avoidance considerations.
3. Specify the constraints related to the project as a whole, or individual stages of the project, and see the recommended methods.
4. View a summary of the recommendations for the project in the form of a usability plan including the selection of methods for each stage.

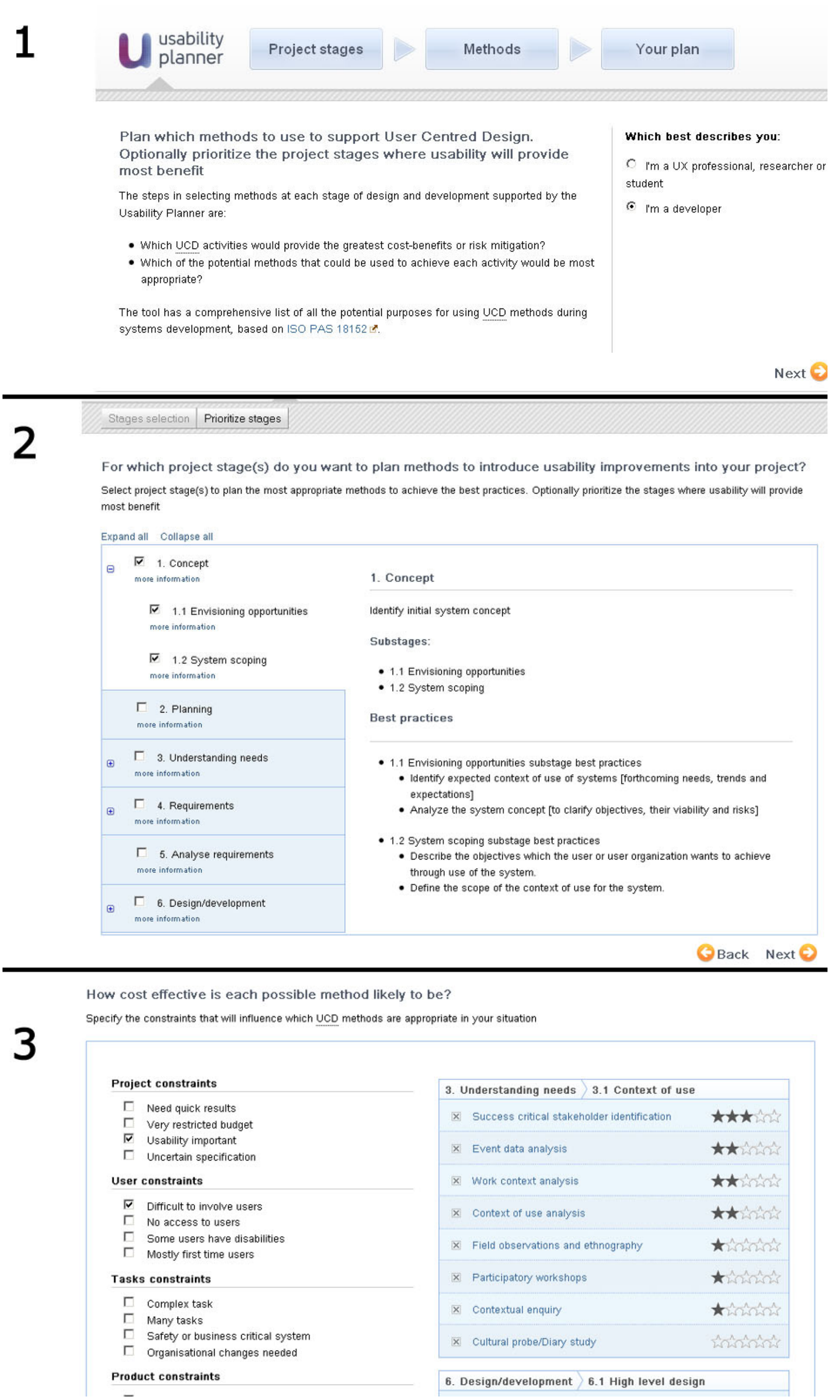


Fig. 1. Three first steps in the usage of the tool for method selection

3 Preliminary Evaluation and Conclusions

When a preliminary prototype [2] was demonstrated to participants at the NordiCHI 2010 conference, all 11 people who tried using the tool were positive about its potential value. One continuing challenge has been to find a sequence that the different types of users find logical for the inter-related decisions that have to be made. Decisions on inter-stage prioritization have been made optional, since some users will not need to address such an issue, and the support that may provided for such prioritization is weaker than the support provided for method selection.

While the tool initially embodies the knowledge of the authors, this will be refined by iterative evaluation. Additionally, the content is customizable, so that any organization with UCD expertise may modify the criteria for method selection in the tool to match their own criteria.

The tool will be offered under an open source license (GNU General Public License, version 3 [3]), so that the community can modify it or further extend it.

Initial responses to the value of the first complete prototype of the tool have been very positive. The next steps include:

1. Refining the set of methods and criteria, and the internal weighting, to find out whether the tool can consistently make similar recommendations to experts.
2. More formal evaluation of the usefulness and usability of the tool for the intended user groups.
3. Exploring whether the tool can take account of how methods are customised and adapted to project needs.
4. Extending the tool with additional functionalities like saving the current project.

References

1. BCS SIGiST: Usability Guidelines (2006), http://www.testingstandards.co.uk/usability_guidelines.htm
2. Ferre, X., Bevan, N., Escobar, T.A.: UCD Method Selection with Usability Planner. In: Proceedings of NordiCHI 2010 (2010)
3. GNU. General Public License version 3 (2007), http://www.gnu.org/licenses/gpl.html
4. ISO/PAS 18152: A specification for the process assessment of human-system issues (2003)
5. ISO/TR 16982: Usability methods supporting human-centred design (2002)
6. Seffah, A., Metzker, E.: The obstacles and myths of usability and software engineering. Communications of the ACM 47(12), 71–76 (2004)
7. Usability Planner tool (2011), http://usabilityplanner.org/

User Experience Specification through Quality Attributes

Llúcia Masip, Marta Oliva, and Toni Granollers

Dept. de Informática e Ingeniería Industrial,
Universidad de Lleida,
25001, Lleida
{lluciamaar,oliva,tonig}@diei.udl.cat

Abstract. The concept of user experience includes different facets which have still not reached a consensus. On the other hand, the ISO/IEC 25010:2011 standard shows a structured quality model which permits us to obtain quality systems and software. The main aim is the specification of user experience based on the facets which are implicitly considered in the standard.

Keywords: User experience, facets, ISO, quality systems and software.

1 Introduction

Recently, the User eXperience concept (UX) has become more used than usability in the Human Computer Interaction (HCI) discipline and SQUARE (Systems and Software Quality Requirements and Evaluation) [1] is the standard that defines the system quality or the software quality. Nevertheless, HCI practitioners still have not come to an agreement about defining UX. Also, the standard attributes which researchers use for analyzing the UX in a specific interactive system do not exist.

For these reasons, the main aim of this research is the specification of UX based on the facets that the ISO/IEC 2010:2011 considers according to its attributes.

2 User Experience

The ISO DIS 9241-210:2008 [2] standard provides one of the most distinguished definitions for UX. But the following five definitions are also used considerably [3] [4] [5] [6] y [7], which were collected by E. L Law et al. in [8].

Despite the fact that the definitions presented are valid in specific contexts, they do not include aspects which should be considered when evaluating UX. In some definitions such as in [8] and [7], the interaction context is not so clear. In [3], the main topic is concerned in company aspects. Other definitions do not refer to facets such as accessibility [4], cross-cultural [5] or adaptability [6]. So, we propose the next definition of UX which covers these aspects: *"User experience deals with all facts, internal as well as external facts of the user and interactive systems, which causes any feeling in the user who uses the interactive system in a specific context of use."*

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 656–660, 2011.

2.1 Facets and Concepts Involved in the User Experience Definition

The facets considered in UX are still not agreed on in the scientific community or in any type of organization for standardization. Related works in this area are [9] [4]. In addition, there is other concepts which can form the UX: accessibility [10], emotional [11], communicability [12], cross-cultural [13], plasticity [14], playability [15] and dependability [16], among others. Thus, one or another facet is used according to the author and their needs in the design or evaluation process.

Our goal is not choosing the specific facets, but it is preparing the most complete set of UX facets which allows us to carry out this project.

The other goal is to determine the meaning of the most used words in the UX area. According to Oxford's dictionary (http://oxforddictionaries.com/): **Property**: an attribute, quality, or characteristic of something. **Facet**: a particular aspect or feature of something. **Dimension**: a measurable extent of a particular kind, such as length, breadth, depth, or height. **Feature**: a distinctive attribute or aspect of something.

Category: a class or division of people or things regarded as having particular shared characteristics. **Attribute**: a quality or feature regarded as a characteristic or inherent part of someone or something.

We are going to use these terms: *Facet* for determining all UX components; *attributes* for all features, subfeatures and attributes in the standard, and, finally, we will use *dimension* for measurable quality attributes considered in ISO 2502n.

3 User Experience and ISO Standard

The facets which are implicitly considered in the standard are specified in Table 1.

Table 1. UX facets considered in every ISO attribute.

		ISO/IEC 25010 : 2011	UX facets
4.1 Quality in use		**4.1.1 Effectiveness**	Usability, Playability, Useful
		4.1.2 Efficiency	Usability, Playability
	4.1.3 Satisfaction	4.1.3.1 Usefulness	Useful
		4.1.3.2 Trust	Emotional, Playability, Desirable
		4.1.3.3 Pleasure	Emotional, Playability, Desirable
		4.1.3.4 Comfort	Emotional, Playability, Desirable
	4.1.4 Freedom for risk	4.1.4.1 Economic risk mitigation	Dependability
		4.1.4.2 Health and safety risk mitigation	Dependability
		4.1.4.3 Environmental risk mitigation	Dependability
	4.1.5 Context coverage	4.1.5.1 Context completeness	Usability
		4.1.5.2 Flexibility	Usability and Accessibility

Table 1. (*Continued*)

	ISO/IEC 25010 : 2011		UX facets
4.2 Product quality	**4.2.1 Functional suitability**	4.2.1.1 Functional completeness	Useful
		4.2.1.2 Functional correctness	Accessibility, Playability
		4.2.1.3 Functional appropriateness	Accessibility, Playability
	4.2.2 Performance efficiency	4.2.2.1 Time behavior	Usability
		4.2.2.2 Resource utilization	Dependability, Accessibility
		4.2.2.3 Capacity	Dependability
	4.2.3 Compatibility	4.2.3.1 Co-existence	Plasticity
		4.2.3.1 Interoperability	Accessibility, Plasticity
	4.2.4 Usability	4.2.4.1 Appropriateness recognizability	Usability, Findable
		4.2.4.2 Learnability	Usability, Playability
		4.2.4.3 Operability	Usability
		4.2.4.4 User error protection	Usability, Playability
		4.2.4.5 User interface aesthetics	Usability, Playability
		4.2.4.6 Accessibility	Accessibility
	4.2.5 Reliability	4.2.5.1 Maturity	Dependability
		4.2.5.2 Availability	Dependability, Accessibility
		4.2.5.3 Fault tolerance	Dependability
		4.2.5.4 Recoverability	Dependability
	4.2.6 Security	4.2.6.1 Confidentiality	Dependability
		4.2.6.2 Integrity	Dependability
		4.2.6.3 Non-repudiation	Dependability
		4.2.6.4 Accountability	Dependability
		4.2.6.5 Authenticity	Dependability
	4.2.7 Maintainability	4.2.7.1 Modularity	Dependability
		4.2.7.2 Reusability	Dependability
		4.2.7.3 Analyzability	Dependability
		4.2.7.4 Modifiability	Dependability, Accessibility
		4.2.7.5 Testability	Dependability
	4.2.8 Portability	4.2.8.1 Adaptability	Accessibility, Plasticity
		4.2.8.2 Installability	Plasticity
		4.2.8.3 Replaceability	Plasticity

4 Conclusions

The first topic that we want to highlight is that there are attributes which are considered in more than one UX facet. So, the non-isolation of UX facets is validated.

In addition, all standard attributes are considered by some UX facets; in fact it causes a direct relation between both facets and standard attributes.

In Table 2 we can see the amount of attributes which we detected in each facet.

Despite the results of this research and as UX experts, we believe that two more facets are needed in the design or evaluation process and when other facets are applied. Both these facets could be worked in a transverse and they are called communicability [12] and cross-cultural [13]. So, we can differentiate two types of facets. The parallel facets (dependability, usability, playability, plasticity, accessibility, emotional, desirable, findable and useful) are those which can be applied in an interactive system in an individual way. And transverse facets are those which could be applied at the same time as when another facet is applied.

Table 2. Amount of considered attributes in each UX facet. UX facets.

UX facets	Amount of attributes
Dependability	19
Usability	10
Playability	10
Accessibility	9
Plasticity	5

UX facets	Amount of attributes
Emotional	3
Desirable	3
Useful	3
Findable	1

Acknowledgements: The work has been partially supported by Spanish Ministry of Science and Innovation through the Open Platform for Multichannel Content Distribution Management research project (TIN2008-06228) and it has been partially supported by the University of Lleida for a pre-doctoral fellowship to Llúcia Masip.

References

1. ISO, International Software Quality Standard, ISO/IEC 25010. Systems and software engineering-Systems and software Quality Requirements and Evaluation (SQuaRE) (2011)
2. ISO DIS 9241-210:2008. Ergonomics of human system interaction - Part 210: Human-centred design for interactive systems (formerly known as 13407). ISO. Switzerland
3. `http://www.nngroup.com/about/userexperience.html` (última visita 4 marzo 2011)
4. Hassenzahl, M., Tractinsky, N.: User Experience - a research agenda (Editorial). Behavior & Information Technology 25(2), 91–97 (2006)
5. Desmet, P.M.A., Hekkert, P.: Framework of product experience. International Journal of Design 1(1), 57–66 (2007)
6. Sward, D., MacArthur, G.: Making user experience a business strategy. In: Law, et al. (eds.) Procs. of the Workshop on Towards a UX Manifesto, Lancaster, UK, pp. 35–40 (2007)
7. `http://www.uxnet.org/` (last view 4 marzo 2011)
8. Lai-Chong, E., Roto, V., Hassenzahl, M., Vermeeren, A.: Kort. J. Understanding, scoping and defining user experience: a survey approach. In: Proceedings of the 27th International Conference on Human Factors in Computing Systems, pp. 719–728. ACM, NY (2009)
9. Morville, P.: Experience design unplugged. In: Kothary, N. (ed.) ACM SIGGRAPH 2005 Web Program (SIGGRAPH 2005). ACM, New York (2005); Article 10
10. W3C (2008). Web Content Accessibility Guidelines 2.0. W3C Candidate Recommendation (April 2008), `http://www.w3.org/TR/WCAG20/`
11. Leitner, G., Hitz, M., Melcher, R.: The Role of Usability in the Design and Evaluation of Dynamic Traffic Displays. In: Holzinger, A. (ed.) USAB 2008. LNCS, vol. 5298, pp. 205–220. Springer, Heidelberg (2008)
12. Prates, R., de Souza, C., Simone, D.J.: Methods and tools: a method for evaluating the communicability of user interfaces. Interactions, 31–38 (2000)
13. Jiang, O., De Bruijn, O., De Angeli, A.: The Perception of Cultural Differences in Online Selfpresentation. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5726, pp. 672–685. Springer, Heidelberg (2009)

14. Thevenin, D., Coutaz, J.: Plasticity of User Interfaces: Framework and Research Agenda. In: Conference on Human-Computer Interaction, INTERACT 1999, vol. 1, pp. 110–117 (1999)
15. González, J.L., Padilla, N., Gutiérrez, F.: From Usability to Playability: Introduction to Player-Centred Video Game Development Process. In: Kurosu, M. (ed.) HCD 2009. LNCS, vol. 5619, pp. 65–74. Springer, Heidelberg (2009)
16. Avizienis, A., Laprie, J.C., Randell, B., Landwehr, C.: Basic Concepts and Taxonomy of Dependable and Secure Computing. IEEE Trans. Dependable Secur. Comput. 1, 1 (2004)

Using Availability Heuristics in Game Design to Introduce Children to Energy Sufficient Behaviours at Home

Nsemeke Ukpong[1], Privender Saini[2], and Abdullah Al Mahmud[1]

[1] Eindhoven University of Technology, Den Dolech 2,
5612 AZ Eindhoven, The Netherlands
[2] Philips Research High Tech Campus 34, 5656 AE Eindhoven, The Netherlands
N.I.Ukpong@Student.tue.nl, Privender.Saini@Philips.com,
A.Al-Mahmud@Tue.nl

Abstract. Parents looking to foster more energy sufficient behaviours in their children struggle to have their child maintain these behaviours unsupervised. Our research indicates that in order for the child to maintain these behaviours, s/he must perceive him/herself as an eco conscious individual. We propose that increasing a child's ability to firstly recognise eco-friendly behaviours and secondly, recollect them, is likely to yield a greater affinity for habitual energy sufficient behaviours. This paper describes a first prototype game, whose interface employs availability heuristics and other persuasive design elements to achieve this goal.

1 Introduction

The negative impact of increasing global energy consumption on our environment is becoming alarmingly apparent [1], [2]. It is recognized that the problem has to be tackled and from multiple angles, for example recycling schemes for plastics and glass and energy efficiency labels for electrical appliances [3]. We describe the user-centred design of a dedicated system aimed to foster energy sufficient behaviours at home in children aged between 8 and 10 years. Our user group is inspired by the idea that knowledge and skills practiced in childhood, will lead to eco-friendly decision making in adulthood.

To gauge the desirability of such a system, we conducted an online survey involving 25 parents. The survey outcomes indicated that parents would indeed welcome more energy sufficient behaviors in their children. In addition, it revealed that the challenge lies in getting the children to adopt these behaviors unsupervised. This is thus the focus of the system explained within this paper.

2 Design: Approach and Concept

Once established, human behaviors generally take a long time to change. Literature research into this area indicates that there is a feed-forward loop between attitude and self perception [5]. According to Cornelissen et al. [5], self perception is dependent on the availability heuristic of specific behavior-related information.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 661–664, 2011.

In order to gauge the connection between children's attitude and their knowledge of energy sufficient behaviors five children of ages between 8 and 10 were interviewed. The interviews were conducted at 2 different international primary schools in the Netherlands. During the interviews, a quiet location at the child's primary school was chosen and he or she was asked a set of pre-made questions in a conversational manner. The questions were designed to gather information regarding the relation of the following four key areas: knowledge, activities, communication and attitude, and energy sufficiency.

Our findings showed that the children were aware of direct actions such as switching off electrical appliances such as lamps, televisions and computers. Upon probing, the children were able to identify further energy saving actions such as closing doors or dressing warmly in the winter. We observed the need for probing, and the outcomes as an indication that there were likely to be further actions which children carry out at home but which, due to their indirect relationship with energy, were not readily associated with eco-friendliness. This observation inspired the primary function of the proposed system: to increase the children's ease of retrieving pro-ecological actions.

The envisioned design is an intelligent system that facilitates a game employing role play to challenge children to independently execute character-specific energy sufficient actions during their normal daily routine. These actions are monitored by the system and feedback that is both stimulating and educational is provided in a manner that reinforces the child's ability to associate the action with eco-friendliness. It is our expectation that long term play of this game, will strengthen the child's long term perception of him/herself as an eco-conscious individual.

3 Concept: Implementation and Evaluation

The system described in Fig 1. implements 3 preliminary components identified as being the foundation of the conceptual model of the envisioned game. These components are 'Character Cards', an 'Activity Tree' and a list of 'Alternative Energy Sufficient Behavior' (AESBs). For ease of evaluation, it has been developed as an interactive web based interface. Our objective is not to test the 'real time' properties of the game but rather, if the enforced relationship between the 3 key components is suitable for increasing children's knowledge of AESBs. The relationships between the components are described below:

- AESBs describe different energy sufficient behaviors that children aged between 8 and10 might carry out a specific activity during their daily (weekday) routine. They are presented in text format, in a list of 5. These five comprise of 1 energy-neutral behavior, and 4 energy-impactful behaviors. Of these 4, only 2 are fitting to the persona of a 'Character Card'.
- A 'Character Card' is an item depicting a child aged between 8 and10 who, due to his or her personality, is inclined towards specific AESBs. Of the five AESBs listed, only two correctly match both the activity presented by the 'Activity Tree', as well as the profile of a child on a specific 'Character Card'. The objective of the game is to identify the 2 correct AESBs.

- Activities are mapped to the branches of an 'Activity Tree' visualization. This mapping has the purpose of communicating the impact that different behaviors can have on the tree during a specific routine activity such as breakfasting. The tree communicates its health through its number of leaves and the brightness of its leaves.

To play the game, the child selects either to try an evening or a morning activity. Upon doing so, the timer starts and s/he is presented with the elements in stage 2 of Fig 1. The child then selects and deselects the AESB as s/he sees fit, pressing the 'Check Guess' button when ready to test the guess. If the guess is correct, the timer stops, all the targeted leaves from dark green to bright green and a congratulatory message appears on the screen. If the guess is incorrect a message appears for a short period, informing the child of this, the child should then try again.

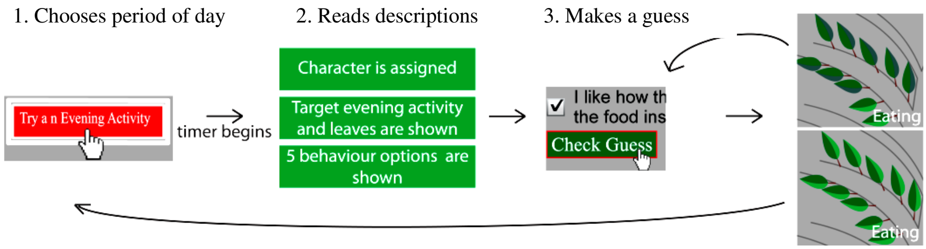


Fig. 1. Example outline of the game sequence supported by the interactive web-based prototype

By varying the combinations between 'Character Card', targeted activity and number of targeted leaves, the game can offer varying levels of difficulty.

3.1 Scenario Evaluation

For the test scenario, 4 children aged between 9 and 11 were provided with a URL link to the game so that they could play it individually in their own homes. The children were instructed to play the game twice: once for the morning routine and once for the evening routine. Prior to playing the game, they were asked to read through a pre-test form presenting two lists of the same 13 energy impacting behaviors. Amongst these 13 were the AESBs presented in the game scenario. For one list, they were asked to select the behaviors which they thought best completed the following sentence: *"When I want to save energy around breakfast time I can…"*and for the second list, the following sentence was used: *"When I want to save energy around relaxing time in the evening I can…"* After playing the game once per period of day, they were required to fill in a post test form involving the same list and activity as the pre test form. Finally, the forms were sent via email.

Each behavior had a weighted value of 0, 1 or 2, depending on its suitability to the headline sentence. This weight was used to calculate the scores achieved by the children. The results showed a lack of improvement by 18 of the behaviors. Of this 18, 9 were negative results- i.e. the child selected the behavior in the pre test form but not in the post test form. The other 9 showed no change. Six out of the 26 available behaviors showed positive improvement. Four out of this 6 belonged to the morning activity. Child 1,2 and 4 showed an increase in their scores of percentages 12, 3 and 3 respectively. Child 3 showed a decrease of 3%.

The results showed that there was a marginal increase in the children's familiarity with AESB: with 3 out of the 4 children being able to identify AESBs more correctly than before. The children's remarks regarding the textual feedback in particular indicated that they understood the mapping between successful guesses and unsuccessful guesses. This indicated that the game had been successful in introducing the participants to AESB.

4 Discussion and Conclusions

We found that with a short test, we were able to increase the children's ability to identify more energy efficient behaviors than before. This is a promising finding, as it gives confidence in the potential impacts such a game could have.

This paper has made one major contribution towards the design of educational interactive systems for initiating behavior changes in children. We have shown with a first prototype game designed to present alternatives to habitual behaviors to positively influence a child's affinity to energy sufficient behaviors.

Further work could focus on extending the list of eco-friendly behaviors children could perform, and even have lists tailored to certain ages. In addition, more thought could be put in the 'Character Card's to appeal to children of various ages. Finally, a more robust and extended set-up would allow us to conduct a more thorough test regarding the potential of the game to add knowledge, and ultimately change behavior.

References

1. Ehrlich, P.R., Holdren, J.P.: Impact of Population Growth. Science 171, 1212–1217 (1971)
2. World Energy Outlook: Fact Sheet (2009), `http://www.worldenergyoutlook.org/docs/weo2009/fact_sheets_WEO_2009.pdf`
3. Environmental Protection Agency: Energy Conservation (1997), `http://www.epa.gov/reg5rcra/wptdiv/p2pages/energy.pdf`
4. Cornelissen, G., Pandelaere, M., Warlop, L.: Cueing Common Ecological Behaviors to Increase Environmental Attitudes. In: IJsselsteijn, W.A., de Kort, Y.A.W., Midden, C., Eggen, B., van den Hoven, E. (eds.) PERSUASIVE 2006. LNCS, vol. 3962, pp. 39–44. Springer, Heidelberg (2006)

UsiXML Extension for Awareness Support

Jose Figueroa-Martinez[1], Francisco L. Gutiérrez Vela[1],
Víctor López-Jaquero[2], and Pascual González[2]

[1] University of Granada,
C/ Cuesta del Hospicio s/n – 18071,
Granada, Spain
[2] University of Castilla-La Mancha,
CP 02071,
Albacete, Spain
{jfigueroa,fgutierr}@ugr.es, victor@dsi.uclm.es,
Pascual.Gonzalez@uclm.es

Abstract. Awareness support in MDA technologies is virtually nonexistent. Furthermore, until recently there was no conceptual model suitable for representing Awareness support in model based architectures. Here, we introduce an extension of UsiXML user interface description language to support Awareness as an information requirement. UsiXML aims to describe multi-context and multimodal user interfaces. The model-based approach of UsiXML makes it a good candidate for integrating *Awareness Support* from the requirements phase to the final user interfaces. It enables Awareness requirements to be traced from the final user interfaces to the tasks and domain entities that generate them, allowing developers to maintain and validate all the Awareness mechanisms provided by the system. This leads not only to a better quality of system developed, but also an organized and traceable development of Awareness mechanisms, easier maintenance and improved user interaction.

Keywords: Awareness, model-based user interface development, requirements.

1 Introduction

The interest in model-based development of user interfaces [1] has been expressed through the creation of a W3C Incubator Group [2], which takes into account the standardization of model-based user interface design. One of the most active User Interface Description Languages (UIDL), based on the available tools and the community that supports it, is UsiXML [3], which is a XML based markup language used to describe multi-context and multi-platform user interfaces at different levels of abstraction.

UsiXML, as well as other UIDL, does not support the specification of awareness requirements nor the awareness itself, even though awareness features should be part of the specification of any collaborative system.

In general, Awareness means “the knowledge of what is going on” [4]. Awareness support as part of modern model-based development tools is still limited [5], mainly because the gap of development-oriented conceptual models.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 665–668, 2011.

UsiXML is based on the Cameleon Reference Framework [6], which defines UI development steps for multi-context interactive applications. The development steps are Task & Concepts (T&C), Abstract User Interface (AUI), Concrete User Interface (CUI) and Final User Interface (FUI), which represents the operational UI.

The objective of this work is to extend UsiXML step by step, introducing Awareness as a new requirement that can be defined, linked, propagated and validated at all stages of the methodology.

2 Awareness Support in UsiXML

We have created several models which will provide the scaffolding for awareness support in UsiXML. The standard transformation process must be changed in order to generate the new mappings that make awareness support traceable.

First, we demonstrate how Awareness is represented. Next, we show how Awareness is included in a domain, in order to obtain a usable source of data. We subsequently describe how to use Awareness data and how to include it in the software development process.

Generic Awareness Representation

According to Endsley's description of awareness [4], an Awareness of "something" (the existence of a resource, the location of some entity, the duration of a process, etc.) is defined by a set of information elements. "Elements" signify something that the observer can receive and understand, such as location, height, weight, size, speed, etc. These concrete elements are features of the observed entity.

There may be other elements that are created by the understanding of the observer, which we call composite elements. The *composite element* is defined as a composition of *concrete elements* and/or other *composite elements,* joined by a *compositing function*. Furthermore, it may also be possible to obtain projections of the value of some elements in the near future. We call this Awareness Projections (Fig. 1a).

This Awareness Abstract Representation (AAR) can be useful for sharing and reusing generic models of awareness types that are already known. However, for concrete Awareness support a domain-specific awareness representation is required.

Domain specific Awareness representation

In order to define specific awareness types for the entities of a particular domain, these awareness types must be linked to domain class attributes, in order to create usable awareness data sources. We call this combination an Awareness Concrete Type (ACT) and it represents the way to define the supported awareness in a system (Fig. 1a). We should be able to define many ACTs, some of them from the same AAR. For example, *Group Location Awareness*, *User Location Awareness*, *Pointer Location Awareness*, are all conceptually equal to *Location Awareness*.

In order to define an ACT we must create a structure (similar to the AAR that the type comes from) which is linked to the observed entity and which has the concrete elements (and composite elements) linked to the corresponding attributes (*composeAttribute*) of the observed entity, as shown in Fig. 1a. It is necessary to extend the Domain Model with the entities shown in Fig. 2a in order to define composed attributes.

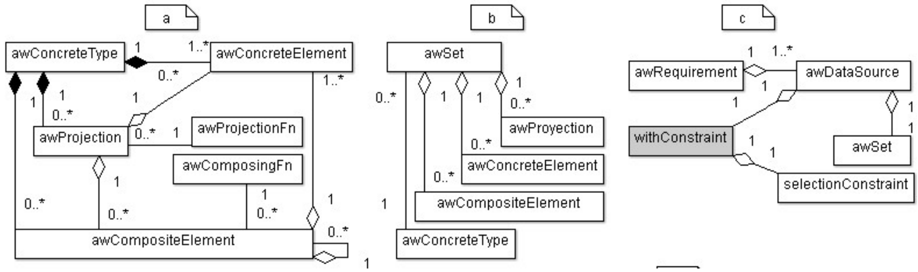


Fig. 1. a) ACT model. b) Awareness Set model. c) Awareness Data Source Model.

Awareness Requirement representation

Once awareness types have been defined for the domain, the next step is to use them. For this, some UsiXML model changes are required to help deal with issues such as privacy and disruption. Privacy means that awareness data is only given to the users that need it. Accordingly, a new entity is added to the Context Model (Fig. 2b) to deal with this Runtime Condition (RC). To reduce disruption problems, the system should provide users with only the awareness data they need (Fig. 1 b & c).

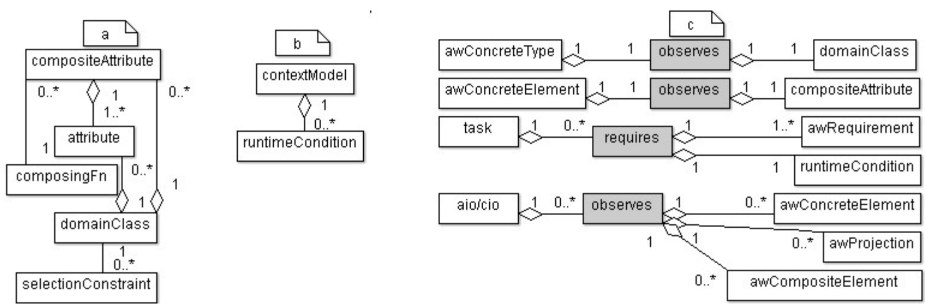


Fig. 2. a) Extension to Domain Model. b) Extension to Context Model. c) Added mappings.

The system should provide all the awareness information required by any user for performing a task. This means that during a task, users must receive the updated data they require, which can be different depending on the runtime context. Our proposal is to add Awareness Requirements (AR) to a task (Fig. 2c) by using mappings [7]. An AR is the representation of an awareness information requirement for a specific case or context (RC) which accesses a selected group of instances from the awareness data source (*awDataSource*, *awSet*). Fig. 1 b & c.

3 Conclusions and Future Work

This work presents a model-based approach for developing awareness supported muti-platform and multi-modal UIs based on the UsiXML methodology. The novelty of this proposal is that Awareness is included as an information requirement and is managed at all stages of the methodology, allowing reusability, traceability, verifiability as well as the ability to be processed by software.

The logical separation of awareness as a requirement and user interfaces improves the quality and maintainability of both, and creates new ways to improve the development process.

The *awareness support* defined as a requirement model opens the door to other forms of abstractions. The knowledge stored in the models can be used in novel ways and offer more advantages to developers and users. Furthermore, some important problems like privacy and disruption related to the awareness support are tackled by using the "runtime conditions" along with the Awareness Requirements, and by using "selection constraints" that clearly select the awareness data sources.

As future work, we plan to improve the transformation process that manages the awareness models through the different steps of the UsiXML software development process. We also plan to include awareness models in other User Interface Definition Languages and development methodologies.

Acknowledgements. This research is financed by the Ministry of Science and Innovation, Spain, as part of DESACO Project (TIN2008-06596) and the Mexican National Council of Science and Technology CONACYT.

References

1. Puerta, A.: A Model-Based Interface Development. IEEE Software 14(4), 40–47 (1997)
2. Cantera, J.M.: Model-Based UI XG Final Report. W3C Incubator Group Report, May 04 (2010), `http://www.w3.org/2005/Incubator/model-based-ui/XGR-mbui/`
3. Limbourg, Q., Vanderdonckt, J., Michotte, B., Bouillon, L., López Jaquero, V.: UsiXML: a Language Supporting Multi-Path Development of User Interfaces. In: Feige, U., Roth, J. (eds.) DSV-IS 2004 and EHCI 2004. LNCS, vol. 3425, pp. 200–220. Springer, Heidelberg (2005)
4. Endsley, M.R.: Towards a Theory of Situation Awareness in Dynamic Systems. Human Factors 37(1), 32–64 (1995)
5. Figueroa Martinez, J., Gutierrez Vela, F.L., Collazos, C.A.: Awareness Models for the Development of Ubiquitous Systems. In: Augusto, J.C., Corchado, J.M., Novais, P., Analide, C. (eds.) ISAmI 2010. Advances in Intelligent and Soft Computing, vol. 72, pp. 237–245. Springer, Heidelberg (2010)
6. Calvary, G., Coutaz, J., Thevenin, D., Limbourg, Q., Bouillon, L., Vanderdonckt, J.: A Unifying Reference Framework for Multi-Target User Interfaces. Interacting with Computers 15(3), 289–308 (2003)
7. Montero, F., López Jaquero, V., Vanderdonckt, J., González, P., Lozano, M.D.: Solving the Mapping Problem in User Interface Design by Seamless Integration in IdealXML. In: Gilroy, S.W., Harrison, M.D. (eds.) DSV-IS 2005. LNCS, vol. 3941, pp. 302–9743. Springer, Heidelberg (2006) ISSN: 0302-9743

Web Accessibility Requirements for Media Players

María González, Lourdes Moreno, Paloma Martínez, and Ana Iglesias

LaBDA Group, Computer Science Department, Universidad Carlos III de Madrid,
Avda. Universidad 30, 28911 Leganés, Madrid, Spain
{mgonza1,lmoreno,pmf,aiglesia}@inf.uc3m.es

Abstract. Video Content continues to strenghten on the Web, because of that fact, it is neccessary to include video content with suitable accessibility requirements to be used by all people. User Agent (such as player, browser), also have to include accessibility requirements. In this paper a set of basic guidelines is included for professionals who want to embed video players on their Web.

Keywords: Web accessibility, user agent, media player, standard, evaluation.

1 Introduction

The Web is being overflowed with multimedia content, like video. This will keep on happening into the future [1]. Not only multimedia content has to be accessible through standards like the Web Content Accessibility Guideline (WCAG) [1] of Web Accessibility Initiative (WAI) [3] among others, but it also has to satisfy accessibility requirements, which have independent components, for example those that are referred to the user agent such as media players, browser, etc.

Sometimes, accessibility problems appear because of accessibility barriers within the media player, such as browsers, assistive technologies or even user's inability, for example, when a user tries to watch a video, but do not have the appropriate software and do not how to install it.

To avoid these problems, media players have to be developed in agreement with User Agent Accessibility Guidelines (UAAG) [4] of WAI, as well as, having into account different opinions based on studies to offer universal access solutions for the user.

The remaining of the paper is organized as follows: Section 2 discusses different accessibility standards for media players and provides technological solutions. In Section 3, accessibility guidelines for media players are presented. Finally, Section 4 presents the concluding remarks and future works.

2 Background

In this section, standards for Web accessibility, regulations about how to standardize accessibility in software and good practices guidelines will be indicated, as well as new contributions that are provided by the new standard HTML5.

The UAAG of WAI is an international standard that explains how to make user agent accessible for people with disabilities and how to increase accessibility on Web

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 669–674, 2011.

content. Nowadays, this standard has two versions, UAAG 1.0 [5] (the reference version) and UAAG 2.0 [6] (the draft version). Last version gives support to WCAG 2.0 [7], being this one used to make the accessibility guidelines.

Another standard is ISO 9241-171:2008 [8], Ergonomics of human-system interaction, this standard provides a group of guidance on software accessibility.

The new standard HTML5 [9] allows to play videos without installing plug-ins through the new labels such as <video> and <audio>. However, there are some problems due to the fact that it does not support all Web user agent completely in the keyboard access or through screen readers to the provided controls. On the other hand, the current draft version of HTML5 does not give support to include subtitles and audio description, so in order to become the official recommendation, it is necessary to include these requirements to satisfy UAAG 2.0 and WCAG 2.0 [10].

As well as these standards, there are works [11][12][13] and institutions [14][15][16] that provide good practices guidelines according to accessible multimedia content on the Web.

3 Accessibility Guidelines for Web Media Players

With the knowledge that we obtain after carrying out a study of WAI standards such as WCAG 2.0, the agile evaluation method based on UAAG 2.0 as well as the study of others standards and relative works, following, we present as result, a help documentation to guide web professional. The documentation includes a set of basic guidelines and a group of elements to be introduced on a media player interface so a video is accessible on a Web media player. This documentation is elaborated to help Web professionals.

1) A media player has to provide different alternatives for audiovisual information next to the video:
    a. Captions (video subtitles). As it is shown in Figure 1.

Fig. 1. *BBC iPlayer's* screenshot showing subtitles

b. Audio description.
c. Others (sing language, transcription, extended audio description, etc.).

2) Complete access to all the features has to be guaranteed, as well as towards the mouse:
   a. By keyboard.
   b. Through assistive technology (like screen reader).
3) Provide help and documentation about the features of media player accessibility on the user interface, which reports to the user about the availability of those features, as well as the information about their purpose and use.
4) Provide a keyboard focus cursor that shows visually what element has the focus on the user interface; it is also important that the media player has a text cursor to show the focus within a text element. Restore the state when the state focus is recovered.

According to the previous guidelines, elements on the user interface have to be included like the following controls. These controls are divided in two groups: basic controls and additional controls that are necessary to obtain an accessible media player.

In the first group (basic controls) are included:

- Controls that allow users to play or to stop the video.
- Controls that allow users to resize the viewports.
- Controls that allow users to adjust the volume.

Among additional controls, we introduce the following ones:

- Controls that allow users to enable or to disable subtitles. As it is shown in Figure 2, there is a button called ‘CC’ (Closed Caption) that provides this functionality.

Fig. 2. *YouTube’s* screenshot that shows CC button

- Controls that allow users to enable or to disable the audio description.
- Controls that allow users to search within reproduction subtitles, as it is shown in Figure 3.

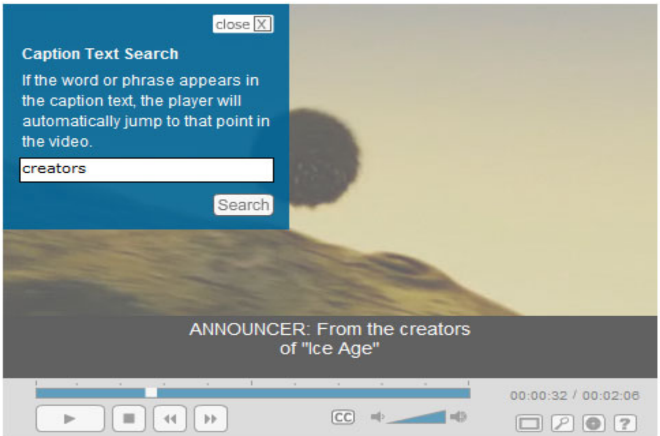


Fig. 3. *CCplayer's* screenshot that shows a caption search

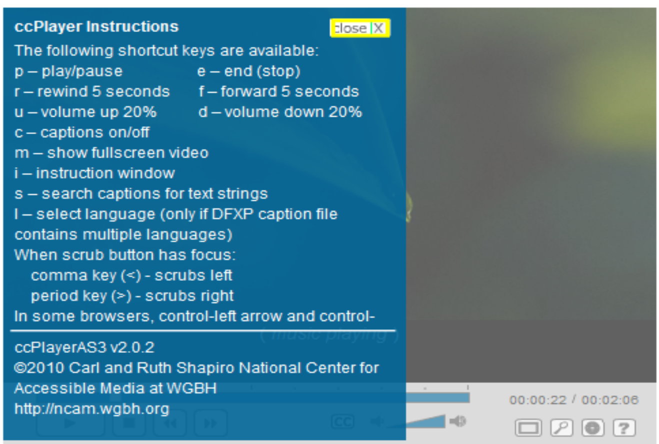


Fig. 4. *CCPlayer's* screenshot that shows a command's menu

- Controls that allow users to forward or to delay seconds within a reproduction.
- Controls that allow users to change the size, font or color of the text.
- Controls that allow users access to help documentation which reports keyboard shortcuts, as it is shown in Figure 4.

5) According to other important requirements concerning accessibility close to usability, the following requirements have been found:

a) It is very important to maintain accessibility features that are configured by the users in the following sessions and that those features only change when the user want.
b) Controls that allow users to enable or to disable and to adjust accessibility features have to be easy to find and they have to be operable.
c) Provide information that allows the user to know the keyboard shortcuts that can be used in the media player, such as the help menu shows in Figure 4.
d) Allow navigating through the content without enabling any controls.
e) Has to be easy so the user can move through the menus, submenus and list through different combinations of keyboards or direct keyboard commands.
f) Allow users to set their preferences to configure the keyboard shortcuts.

4 Conclusions

Although Web multimedia content has grown, this growth is not comparable to the Web accessibility growth, because it is not accessible enough yet. Due to that fact, it is very important that the media players and their access through the Web are accessible.

This paper was elaborated to help Web professionals to create or to decide what media player has to be included on their Website to guarantee the same access for all and accessible content is provided. Because of this, a set of accessibility basic guidelines that have to be satisfied have been introduced to obtain an accessible media player. Besides, an agile method of evaluation is introduced, which was used previously to evaluate some media players, that help evaluators to prove if a media player labeled as accessible actually is.

Acknowledgments. This research work is supported by the Research Network MAVIR (S2009/TIC-1542 (see www.mavir.net/)), and GEMMA (TSI-020302-2010-141) and SAGAS (TSI-020100-2010-184) research projects.

References

1. Cisco Visual Networking Index, http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481360_ns827_Networking_Solutions_White_Paper.html
2. W3C, WAI, Web Content Accessibility Guidelines, WCAG (2010), http://www.w3.org/WAI/intro/wcag.php
3. W3C, Web Accessibility Initiative, WAI (2010), http://www.w3.org/WAI/
4. W3C, WAI, User Agent Accessibility Guidelines, UAAG (2010), http://www.w3.org/WAI/intro/uaag.php
5. W3C, User Agent Accessibility Guidelines 1.0 (2002), http://www.w3.org/TR/WAI-USERAGENT/

6. W3C, User Agent Accessibility Guidelines 2.0 (2010), http://www.w3.org/TR/UAAG20/
7. W3C, Web Content Accessibility Guidelines 2.0 (2008), http://www.w3.org/TR/WCAG20/
8. ISO, International Organization for Standardization (2008), http://www.iso.org
9. W3C, HTML5, A vocabulary and associated APIs for HTML and XHTML, http://www.w3.org/TR/2009/WD-html5-20090423/
10. Moreno, L., Martínez, P., Iglesias, A., González-García, M.: HTML5 support for an accessible user-video-interaction on the Web. In: 13th IFIP TC13 Conference on Human-Computer Interaction, INTERACT 2011, Lisbon, Portugal (September 2011) (in press)
11. WebAIM, Media Player Accessibility (2010), http://webaim.org/techniques/captions/mediaplayers/
12. Clark, J.: Accessibility, Design and Writing (2010), http://joeclark.org/
13. Moreno, L., Martínez, P., Ruiz-Mezcua, B.: Guías metodológicas para contenidos multimedia accesibles en la Web. In: X Congreso Internacional de Interacción Persona-Ordenador, Interacción 2009, Barcelona, España (September 2009) ISBN: 13:978-84-692
14. Alfred, P.: Sloan Foundation (2008), http://www.sloan.org/
15. NCAM, National Center for Accessible Media (2009), http://ncam.wgbh.org/
16. Stanford Captioning, http://captioning.stanford.edu/

Christian Doppler Laboratory: Contextual Interfaces

David Wilfinger, Alexander Meschtscherjakov, Astrid Weiss, and Manfred Tscheligi

Christian Doppler Laboratory "Contextual Interfaces", ICT&S Center, University of Salzburg, Sigmund Haffner Gasse 18, 5020 Salzburg, Austria
firstname.lastname@sbg.ac.at

Abstract. The Christian Doppler Laboratory for Contextual Interfaces is a cooperative research lab dealing with interaction designs of contextual interfaces in the automotive area and the cleanroom of a semiconductor factory. This paper describes the research approach and example activities conducted in the laboratory.

1 Introduction

Human-Computer Interaction (HCI) is dependent on the concrete characteristics of the context, in which the interaction takes place. To design interfaces for a specific usage context with the aim of improving user experience (UX), it is necessary to identify relevant context factors and understand the particular environment in a holistic way. For that purpose, we established a laboratory with the aim of investigating contextual interfaces on a theoretical level, as well as applying these abstract results into two particular contexts: the factory and the car. In comparison to traditional environments these contexts have been rather neglected in HCI research. The so-called Christian Doppler Laboratory for Contextual Interfaces is located at the Center for Advanced Studies and Research on Information and Communication Technologies (ICT&S) at the University of Salzburg, Austria. The laboratory is structured into three main areas: a *basic research area*, a *context factory area*, and a *context car area*. This paper gives a short overview of the organizational entities of the laboratory, as well as some example results.

2 Research Approaches and Results

In the focus of ongoing activities in the laboratory is the understanding of the two interaction contexts in terms of related UX and contextual influences, as well as a methodological framework to investigate these particular contexts. In the following paragraphs the research areas are introduced with example results.

The *basic research area* deals with theoretical challenges of context, UX and related methodology. It delivers basic building blocks for contextual UX research and guides their application in the car and the factory context. Within basic research activities we work on UX and context models extending existing approaches for their use in both the car and factory context. These models are informed by basic research studies. As an example we developed car experience characters within the automotive context such as “car as a tool” and “car for self-appreciation” by conducting a content

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 675–676, 2011.

analysis of posts in an online car forum. As a second part of basic research, methods to investigate contextual interactions (such as probing and experience sampling) were extended and applied for the specific context of the factory (e.g. cleanroom restrictions) and the car (e.g. safety related issues).

The *context factory* research is a cooperation between the laboratory and Infineon Technologies Austria AG. It specifically addresses the cleanroom of a semiconductor factory from a user-centered perspective. In the focus of the factory research is the optimization of the workplace for operators (and subsequently an increase of productivity) by introducing new interfaces, such as ambient technology or collaborative human-robot interfaces into this interaction space. All activities are based on a thorough understanding of the context factory and its situated user experiences. To achieve this, we applied qualitative methods in the cleanroom in cooperation with our industrial partner. Ethnographic studies were conducted investigating, equipments, working routines, and social structure of operators. Other studies, including contextual inquiries aimed at understanding UX related to equipment maintenance, occurring context factors in the clean room, the effect of error messages on the workers, and operators' attitudes to robots.

The *context car* research is a cooperation with AUDIO MOBIL Elektronik GmbH, a tier 1 supplier for the automotive industry. Within the research activities in this context, the car is divided into three design spaces that are of interest for user-centered design: the driver space, the front seat passenger space, and the rear seat passenger space. Together with our industrial partner we aim at understanding the car as an interaction environment combining users in all three spaces. Research in context car includes the influence of particular context factors on drivers' UX. Widening the focus of car HMI research, an ethnographic study aimed at collaboration between drivers and front seat passengers, investigating where assistance can be supported or taken over by technology. Additionally we used cultural probes to understand the usage of a rear seat area by families. Similar to the factory context, interaction studies aimed at understanding UX that evolves out of the usage of a certain system. For that purpose we investigated a multifunctional steering wheel including a touch screen on the steering wheel, cockpit interaction modalities and eco-friendly driver interfaces.

3 Conclusions and Future Work

This work gives a very short overview on example activities within a research laboratory planned for seven years. Due to the organization of the laboratory we were able to conduct both basic and applied research in the car and factory context. The development of theoretical approaches within the basic module supported the research conducted in the two contexts. Within the next years we will develop and study alternative interaction designs for both the car and the factory. The financial support by the Federal Ministry of Economy, Family and Youth and the National Foundation for Research, Technology and Development is gratefully acknowledged.

Interaction Modeling at PROS Research Center

José Ignacio Panach, Nathalie Aquino, and Oscar Pastor

Research Center on Software Production Methods (PROS),
Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain
{jpanach,naquino,opastor}@pros.upv.es
http://www.pros.upv.es

Abstract. This paper describes how the PROS Research Center deals with interaction in the context of a model-driven approach for the development of information systems. Interaction is specified in a conceptual model together with the structure and behavior of the system. Major achievements and current research challenges of PROS in the field of interaction modeling are presented.

1 Introduction

On the one hand, the software engineering community has provided sound models to represent the structure and behavior of software applications. The Entity-Relationship Model or UML Class Diagrams are examples of structural models, while Sequence or Activity UML Diagrams are examples of behavioral models. On the other hand, the human-computer interaction community has provided several proposals for interaction modeling, but none of them has become widely known or used. Even today, the identification of the conceptual primitives that allow interaction to be properly modeled can be considered a research challenge.

The PROS Research Center (Research Center on Software Production Methods) of the Universitat Politècnica de València, Spain, is devoted to improve traditional software production methods by providing model-driven processes to develop quality software. With this goal in mind, PROS has defined an object-oriented and model-driven development method, called OO-Method [5], which automatically generates software systems from conceptual models. PROS recognizes the importance of the interactive perspective and, therefore, sustains that the conceptual model of an information system cannot be considered to be complete after just specifying its structure and behavior. It is also necessary to specify how end-users will interact with the system. Hence, PROS proposes a holistic conceptual model in which interaction is modeled together with structural and behavioral models. In this holistic conceptual model, interaction with the end-user is represented in a Presentation Model.

2 Achievements: The Presentation Model of OO-Method

The Presentation Model [3] of OO-Method provides conceptual primitives to represent the interaction with a software application. A set of *elementary patterns* is provided to build *interaction units*. There are three basic types of interaction units that

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 677–678, 2011.

can be used to represent different interactive scenarios: 1) the execution of a method, 2) the list of the population of a class, and 3) the visualization of details of a specific object. These basic interaction units can be combined to build more complex ones. Furthermore, a *hierarchical action tree* is provided to organize the access to interaction units, defining the menu of the application.

OO-Method and its Presentation Model are supported by a commercial software suite named OlivaNOVA, which was developed by CARE Technologies (http://www.care-t.com). OlivaNOVA allows the conceptual model to be built and applies subsequent transformations until the final code of a fully functional application (structure, behavior, and interaction) is generated for different computing platforms. The OlivaNOVA tool is currently being used successfully in an industrial environment. Therefore, the Presentation Model of OO-Method can be considered a significant achievement of PROS regarding interaction modeling.

3 Current Research Challenges

Currently, PROS continues devoting efforts to the research line of interaction modeling with the aim of enhancing the expressiveness of the Presentation Model. These efforts can be grouped into two categories: approaches to generate usable systems [4] and approaches to customize user interfaces.

With regard to the first group, a method to include functional usability features [2] in any model-driven approach has been proposed. The method is based on the idea of abstracting the information contained in such usability features in order to include them in a conceptual model. Thanks to this abstraction, a model compiler can automatically generate the code for a specific programming language taking as input the conceptual model. This code supports the functionality of the usability features together with the functionality of the whole system. The proposal supports the existing tendency of addressing the treatment of usability in early stages of the software development process. With regard to the second group, a Transformation Templates [1] approach has been defined in order to diversify the kinds of user interfaces that can be generated. In this way, user interfaces can be customized according to context and user preferences before their automatic generation. This benefits the flexibility of the model-driven approach.

References

1. Aquino, N., Vanderdonckt, J., Pastor, O.: Transformation Templates: Adding Flexibility to Model-Driven Engineering of User Interfaces. In: Proc. of SAC 2010, pp. 1195–1202. ACM Press, New York (2010)
2. Juristo, N., Moreno, A.M., Sánchez, M.I.: Guidelines for Eliciting Usability Functionalities. IEEE Transactions on Software Engineering 33, 744–758 (2007)
3. Molina, P.J., Meliá, S., Pastor, O.: JUST-UI: A User Interface Specification Model. In: Proc. of CADUI 2002, Valenciennes, Francia, pp. 63–74 (2002)
4. Panach, J.I., España, S., Moreno, A., Pastor, O.: Dealing with Usability in Model Transformation Technologies. In: Li, Q., Spaccapietra, S., Yu, E., Olivé, A. (eds.) ER 2008. LNCS, vol. 5231, pp. 498–511. Springer, Heidelberg (2008)
5. Pastor, O., Molina, J.: Model-Driven Architecture in Practice. Springer, Heidelberg (2007)

Overview of the Brazilian Computer Society's Council for Human-Computer Interaction (CEIHC)

Cristiano Maciel[1], Elizabeth Furtado[2], Marco Winckler[3], Milene Silveira[4], and Raquel Prates[5]

[1] Universidade Federal de Mato Grosso (UFMT)
[2] Universidade de Fortaleza (UNIFOR)
[4] Université Paul Sabatier (UPS)
[3] Pontifícia Universidade Católica do Rio Grande do Sul (PUC-RS)
[5] Universidade Federal de Minas Gerais (UFMG)

Abstract. CEIHC is an acronym that stands for Council for Human-Computer Interaction. This council is composed by members of the Brazilian Computer Society (SBC) and its main goal is to promote the field of Human-Computer Interaction in Brazil. In this paper, we detail the missions assigned to the CEIHC and its recent activities. Moreover, challenges for the development of this field and future activities are discussed in this paper.

Keywords: Human-Computer Interaction, HCI research, HCI education.

1 Introduction

The Brazilian Computer Society[1] (SBC) is a scientific organization whose main goal is to promote the technical development of computing at its various disciplines that are represented by a dedicated council. The CEIHC[2] is the SBC's council for all issues related the field of Human-Computer Interaction (HCI). This council is usually composed of 5 members: past and future chairs of the Brazilian Symposium in HCI and one regional position who is elected by the Brazilian HCI community in an assembly organized during that symposium. HCI symposium had its first edition in 1998, and is the main forum for HCI publication and discussion in Brazil since then.

Initially CEIHC was created to assure a good communication channel between the HCI community and SBC, as well as act as steering committee of the HCI symposium. Although supporting the event organization continues to be its main task, due to the growth of the community CEIHC has expanded its focus to support the HCI community in Brazil in other relevant issues, namely: HCI education, HCI research and dissemination of HCI field in Brazil. We next briefly present the CEIHC's effort in each direction and what we see as challenges for the near future.

[1] *Sociedade Brasileira de Computação*, `http://www.sbc.org.br/`
[2] *Comissão Especial de Interação Humano-Computador*, `http://comissoes.sbc.org.br/ce-ihc/`

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 679–680, 2011.

2 CEIHC's Attributions and Activities

The CEIHC, on behalf of the SBC, is responsible for the financial and technical aspects of the Brazilian HCI symposium. Beside this symposium, that in its last edition (2010) had approximately 180 participants, in 2002 CEIHC started supporting the Latin-American Conference on HCI (CLIHC).

Another CEIHC's activity is to develop strategies for promoting HCI education in computer science programs. The community has been invited to discuss this theme at the symposium (a working group in 2006 [2], a panel in 2008 [1] and a workshop in 2010).

Besides that CEIHC is associated with scientific organizations such as IFIP and ACM. CEIHC is responsible for indicating to SBC the Brazilian representative at the IFIP TC 13, and also the BR-CHI (ACM/SIGCHI Local Chapter in Brazil) has its annual meeting during the HCI symposium.

In Brazil there are defined criteria for evaluating research, based on the evaluation of conferences and journals in the field and it has a direct impact on researchers' evaluation and support to their research. Thus, one of the main topics of CEIHC is analyze the criteria established for the HCI field.

Dissemination of HCI in Brazil, another attribution of CEIHC, is done mainly through CEIHC's website[2] and an HCI discussion list (ihc-l@sbc.org.br) managed by CEIHC. Also, in 2009 a special issue on HCI was organized for SBC's bulletin.

3 Achievements and Future Challenges

Among the achievements of CEIHC are the dissemination of the success cases of regular organization of HCI events in Brazil; the promotion of the relationship of the Brazilian community with international associations such as IFIP and ACM/SIGCHI; and, the development of actions and projects according to the SBC Challenges 2006-2016, specially concerning to "Citizens' universal access to knowledge" [3]. The challenges to the future include: continuing to develop strategies for promoting HCI education in computer science programs; make national conferences attractive for research, in order to bring visibility of national research outside; find ways to promote the integration of researchers and practionners of HCI related fields in order to discuss their theories and practices and to apply them for the sustainable development of Brazil; and, promote the dissemination, development and use of HCI.

References

1. de Souza, C.S., Baranauskas, M.C., Prates, R.O., Pimenta, M.S.: HCI in Brazil: lessons learned and new perspectives. In: VIII Brazilian Symposium on Human Factors in Computing Systems, Porto Alegre, pp. 358–359 (2008)
2. Silveira, M.S., Prates, R.O.: Uma Proposta da Comunidade para o Ensino de IHC no Brasil. In: XV Workshop sobre Educação em Computação, Rio de Janeiro, pp. 76–84 (2007)
3. SBC. Grand Challenges in Computer Science. Porto Alegre (2006)

Supporting a Multidisciplinary Digital Media Research Community with GRAND Aspirations

Kellogg S. Booth[1] and Eleni Stroulia[2]

[1] Computer Science, The University of British Columbia, Vancouver, BC V6T 1Z4 Canada
[2] Computing Science, University Alberta, Edmonton, AB T6G 2E8, Canada
ksbooth@cs.ubc.ca, stroulia@ualberta.ca

Abstract. The challenges of managing a nationwide "network of centres of excellence" (NCE) are being explored by GRAND, a Canadian NCE comprising over 350 researchers from technical disciplines, social sciences, humanities and the arts within 34 interwoven projects focusing on all aspects of digital media. A complex web of relationships with funding agencies, private and public partners, and researchers is being managed using a purpose-built web-based platform (the GRAND Forum) that supports communication and collaboration across communities. The Forum explicitly represents multiple roles of individuals within the organization through formal and informal user-centred workflows that reflect both symmetric (peer-to-peer) and asymmetric (hierarchical) organizational structures. We describe the principles of each.

Keywords: Interdisciplinary research collaboration.

Grand challenges such as climate change, energy shortage, the threat of pandemics, and dealing with the accelerating pace of cultural shifts in society are being tackled through large-scale, multidisciplinary research teams. Digital media in its many forms is a tool for these organizations, but is itself a challenge to be tackled. The Graphics, Animation, and New Media network of centres of excellence (GRAND) has multi-year funding from the Canadian federal government for five research themes: (1) new media challenges & opportunities, (2) games & interactive simulation, (3) animation, graphics & imaging, (4) social, legal, economic & cultural perspectives, and (5) enabling technologies & methodologies. These span across disciplines, both public and private organizations, and provincial borders. Each theme has significant intersection with concerns familiar to the HCI community, and the management of the GRAND network faces problems well known in HCI for which are pursuing innovative solutions. We describe two aspects of our approach to these challenges.

1 Management Structures to Promote Multidisciplinary Research

Collaboration is "designed into" GRAND. Each of 34 projects has a project leader and co-leader from different organizations, often from different disciplines. Each of 64 principal network investigators (PNIs) is required to engage with multiple projects (three on average) and every project is expected to contribute to the goals of multiple themes. Projects are defined by a combination of top-down and bottom-up processes.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 681–682, 2011.

Individual PNIs "invest" in projects by apportioning their funding amongst projects in which they engage. Theme leaders look for a balance across projects to advance the shared goals of the network. Funding to PNIs is based on contributions to those goals.

Information about the research is collected using a "report once, use often" strategy to minimize duplication of effort. Research activities, publications, prototypes and research artifacts are described by whoever first reports them and then re-used in subsequent reports. A wiki-like approach to maintaining information allows an initial reporter's information to be updated by others, and linked to all relevant projects, rather than requiring it be siloed into individual project reports. This raises mutual awareness and hopefully leads to serendipitous collaboration across projects.

Following Pascal's dictum "The stream is always purer at its source [Les choses valent toujours mieux dans leur source]," the reporting hierarchy is flattened. Those doing the actual research (most often students and postdocs) provide the initial reports, with supervisors, then project leaders, and finally theme leaders synthesizing summaries in which "raw data" can always be accessed by drilling down.

Two special projects, Media Enabled Organizational Workflow (MEOW) and Network Assessment and Validity for Effective Leadership (NAVEL), will support and reflect on the activities of the GRAND community in a yin-and-yang symbiosis as GRAND evolves through its lifecycle. A set of tools built on the Mediawiki platform is a key component of this organizational management strategy.

2 The GRAND Forum: Scaffolding for Collaboration

To streamline workflows and support communication and collaboration across the network, a web-based Forum is being developed that builds on the principles in the previous section. People in GRAND have roles that evolve over time. Explicit role-based mechanisms manage individual's workflows and maintain a history of those roles (from graduate student to network investigator to industry sponsor). Access control policies associated with roles reflect a variety of concerns, including cross-jurisdictional privacy legislation and IP ownership issues.

Activities such as entering information about a publication are simple single-step workflows. Others, such as reporting project milestones, involve the iterative collaborative editing of multiple pieces of information by multiple contributors, each with different reporting deadlines and degrees of editing authority. Those workflows are too complex to be specified simply at the user-interface level and are too human intensive to be "managed" by traditional workflow engines. We are developing a lightweight event-driven workflow specification and enactment web-service to address this requirement.

GRAND will be evaluated in part by whether its members increase interactions with each other over time. Tracking these relationships is at best tedious and at worst exhausting. We are developing a tagging approach – both interactive through the user interface and automatic through a text-analysis process – so that interactions with artifacts managed by the Forum will enable us to recognize communication channels between people and projects. Audit trails to substantiate information in the Forum and to enable effective use by all our members (subject to the access-control practices above) will be supported by full text and faceted search guided by the network's organizational structures, which are part of the scaffolding underlying the Forum.

The Centre for Internationalization and Usability: Enabling Culture-Centred Design for All

José Abdelnour-Nocera, Andy Smith, John Moore, Cecilia Oyugi, Souleymane Camara, Malte Ressin, Sujan Shresta, and Alison Wiles

School of Computing and Technology, University of West London
St Mary's Road, Ealing. London, W5 5RF
{Jose.Abdelnour-Nocera,Andy.Smith,John.Moore,Cecilia.Oyugi, Souleymane.Camara,Malte.Ressin,Sujan.Shresta, Alison.Wiles}@uwl.ac.uk

Abstract. The Centre for Internationalisation and Usability within the School of Computing and Technology at The University of West London aims to enhance understanding of cultural differences in international software development. A particular focus is the development and usability of ICT products in a global market, both in terms of international software development and economic, community and social development. We host a number of researchers and PhD students working in topics such as usability evaluation and culture, socio-technical participatory design, internationalization attitudes of software engineers, mobile learning and library cognitive design.

Keywords: internationalisation, usability, culture, sociotechnical, participatory design, empirical studies of software engineering.

1 Research Challenges and Agenda

The Centre for Internationalisation and Usability (CIU) aims to support software developers in building systems that meet all the needs of end users globally. A particular focus is concerned with software development in a global market and culture-centred design. Our research strategy is to enhance understanding and knowledge of cultural differences in international software development. There are two research challenges that need to be overcome to achieve this:

a) Computer science and interdisciplinary challenges that need to be addressed in order to provide tools, techniques and methods that can be used by software developers and interaction designers. We research the differing nature of these challenges in both the commercial and international development domains.

b) Institutional and governmental challenges to overcome resistance to change so that new tools, techniques and methods are actually implemented commercially. Implicit here is the requirement to collaborate with academics and practitioners internationally, both in promoting HCI and usability generally and localising the discipline itself, particularly in developing countries.

CIU has been successful in exploiting UK and EU funding streams in delivering its research strategy. We have accessed the EU Asia IT&C programme to support

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 683–684, 2011.

usability research in both India and China. We have been the lead partner in the Sino European Systems Usability Network (SESUN) and the Indo European Systems Usability Partnership [1], helping to develop active and sustainable links between Asia and Europe. This has included collaborating on India's first international HCI conference and co-chairing a major conference in Beijing.

CIU was involved in the EPSRC funded VeSeL project - Village e-science for Life – addressing issues associated with the digital divide in Africa. VeSeL aims to fuse educational and environmental objectives to empower African village communities. Our work focused on socio-technical participatory design and the localization of usability methods [2]. More recently, CIU has begun to explore the use of emerging copyleft hardware technologies and how they might contribute to the fields of mobile learning and ICT for development (ICT4D) [3]. We believe that this open approach provides the opportunity for the community to directly influence the hardware roadmap of a device and thus generate sustainable solutions within markets of developing countries. In the e-learning domain, we are looking at university libraries' usability design from a socio-technical perspective. Our work with commercial usability agencies and clients such as Microsoft, HSBC, Siemens and Intercontinental has allowed us to develop a process model for developing usable cross-cultural websites [4]. Current commercial foci of our research concerns issues in relation to integration of internationalisation approaches within software development methods such as Agile [5] and in global software offshoring [6].

The impact of context and culture poses many challenges that cannot be exposed as a one-off evaluation in technology design. Decisions and actions in product design often result in the emergence of cultural and socio-technical implications and issues. CIU's research is aimed at the early identification of these in order to promote enhanced stakeholder participation and better product usability and user experience.

References

1. Smith, A., Joshi, A., Liu, Z., Bannon, L., Gulliksen, J., Li, C.: Institutionalising HCI in Asia. In: Baranauskas, C., Abascal, J., Barbosa, S.D.J. (eds.) INTERACT 2007. LNCS, vol. 4663, pp. 85–99. Springer, Heidelberg (2007)
2. Camara, S.B., Oyugi, C., Abdelnour-Nocera, J., Smith, A.: Augmenting usability: Cultural elicitation in HCI. In: Katre, D., Orngreen, R., Yammiyavar, P., Clemmensen, T. (eds.) HWID 2009. IFIP AICT, vol. 316, pp. 46–56. Springer, Heidelberg (2010)
3. Shrestha, S., Moore, J., Abdelnour-Nocera, J.: Low-cost hardware for ICT4D: what's right and what's left? Special Issue - IEEE Multidisciplinary Engineering Education Magazine (MEEM) 16(1) (2011)
4. Smith, A., Dunckley, L., French, T., Minocha, S., Chang, Y.: A process model for developing usable cross-cultural websites. Interacting with Computers 16, 63–91 (2004)
5. Ressin, M., Abdelnour-Nocera, J., Smith, A.: Of Code and Context: Collaboration Between Developers and Translators. In: Proceedings of CHASE 2011 at ICSE (2011)
6. Lee, D., Smith, A., Mortimer, M.: Cultural differences affecting quality and productivity in Western / Asian offshore software development. In: Proceedings of India HCI 2011 (2011)

Critical Design :: Is It Just Designers Doing Ethnography or Does It Offer Something More for Interaction Design?

Michael Smyth[1], Chris Speed[2], and Martin Brynskov[3]

[1] Centre for Interaction Design, Edinburgh Napier University, UK
[2] Edinburgh College of Art, Edinburgh, UK
[3] Centre for Digital Urban Living,
Aarhus University, Denmark
m.smyth@napier.ac.uk, c.speed@eca.ac.uk, brynskov@imv.au.dk

Abstract. The panel will discuss the contribution of Critical Design to the field of Interaction Design and reflect on the insights that it provides on interaction.

Keywords: Critical Design & Interaction Design.

1 Introduction

This panel will explore whether Critical Design offers new insight as a method for articulating future interactions. Will its promise deliver or are Interaction Designers destined to experience the same feeling of disappointment that HCI practitioners felt after the novelty of Ethnography began to fade? Will the design fictions of Critical Design mature into design facts or will the demand to solve problems short circuit the method? Indeed, is problem solving *per se* an out-moded way of design thinking in the context of today's multi-layered hybrid society?

2 The Panel :: Critical Design and Interaction Design

The panel will comprise of academic practitioners who seek to inform their thinking through practice. Each panellist will describe their research and, in particular, will focus on the value of Critical Design as it pertains to their work. The presentations will set the tone for the panel and will be designed to engage the audience in a discussion about methods through which to better understand interaction. The aim of the panel will be to address questions, such as:

- how do we extend the requirements horizon beyond existing systems in order to identify emerging themes within the field?
- how do we begin to design concepts when we don't know what that design space will look like, let alone who the user population will be?

It is envisaged that the outcome of the panel session will be an increased awareness of the possible contribution of design to the requirements generation process. In particular, how Critical Design can act as a mechanism to ask questions and make us think.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 685–686, 2011.

Requirements generation has a strong tradition within the field of Human Computer Interaction (HCI). Numerous methods have been adopted from a variety of disciplines and these have been tailored to address the specific demands of HCI. Traditional HCI tends to take a rational, engineering focussed approach, that of problem solving, where insufficiently structured problem statements cause difficulties in arriving at appropriate solutions. Part of the resolution of this situation rests on application of methodologies that attempt to refine and clarify the problem statement itself in order to move towards appropriate systems. Interaction Design, as a more recently developed discipline, takes a more 'designerly' approach to this activity by considering both the problem and the solution in a more fluid and intertwined manner, accepting lack of situational certainty and embracing the contextual and evolutionary nature of designed systems in use [1].

The panel will explore how design can shape strategy and set agendas for future generations of technology and design outcomes. The challenge facing such early stage concept generation is to project forward by tapping into higher level needs and desires that are often not obviously apparent. Critical Design [2] presents design as a catalyst or provocation for thought. It is a strategy for exploring the space that lies tantalisingly beyond the current and the now. By contextualising this approach at the edges of our knowledge, it is possible to use design to create 'design fictions' [3]. Examples of fictions will demonstrate their effect of humanising the future and thereby enabling us to focus on the minutiae of behaviour and the subsequent questions that are revealed through the exposure of our needs, desires, habits, rituals, values and priorities. At the core of Critical Design is the attempt to challenge our assumptions and preconceptions about the role that products and services play in everyday life.

In the introduction to the recent book entitled NonObject [4], Barry Katz described design as a means of surveying the bounds of the believable and pressing against the perimeter of the possible. This characterisation of design as a means of 'cultural research' closely parallels the aspirations of Critical Design. Indeed Lukic [4] views design as a way to probe the emotional space between the human and the artefact and, in a wider sense, a more complete understanding of our object world will provide a means through which we can better understand ourselves. We live in a world where everything seems possible and as a consequence have lost the sense of wonder. Maybe Critical Design is an initial step towards regaining that wonder and better understanding our own condition.

References

1. Dorst, K.: Design problems and design paradoxes. Design Issues 22(3), 4–17 (2006)
2. Dunne, A.: Hertzian Tales – Electronic Products, Aesthetic Experience and Critical Design. RCA/CRD Research Publications, Royal College of Art, London (1999)
3. Bleecker, J., Nova, N.: A synchronicity: Design Fictions for Asynchronous Urban Computing, The Architectural League of New York (2009), `http://www.situatedtechnologies.net`
4. Lukic, B.: NonObject. MIT Press, Cambridge (2011)

Everyone is a Designer, Even Executives!

Jannie Lai and Iram Mirza

Citrix Systems, Inc.,
4988 Great America Parkway,
Santa Clara, CA 95054
{jannie.lai,iram.mirza}@citrix.com

Abstract. This panel includes designers, product managers, and executives from various industries. The discussion focuses on how designers can collaborate effectively with product drivers and executives within their organizations to create a design-driven strategy from concept to implementation.

Keywords: Collaboration, collaboration with executives, partners, culture, vision, design, executives, product managers, designers, hands-on design, CEO, enterprise software, consumer products, start-ups, cloud computing, consultants, design influence, design leadership, design-as-a-strategy (DaaS), design process, best practices, recommendations, compelling design.

1 Introduction

Design is no longer an afterthought at any successful company. As more companies make design a strategic focus, designers influence business directions and executive decisions.

This panel will discuss how designers and executives partner and strategize on how to create a strong design culture within their company. Panelists and audience will explore best practices on topics such as understanding the design and executive roles, collaborating on iterative designs, and executing on the product vision. The panelists represent a spectrum of enterprise and startups experience ranging from design to product development teams.

With our recommendations and suggestions, we hope to help the audience build bridges within their organizations and help executives and product teams understand each other's strengths and perspectives. Following are a few questions that we hope will encourage discussion.

1. When is the best time to engage with the executive team to get their feedback on the product design?

We have worked hands-on with CEOs (Chief Executive Officers) and SVPs (Senior Vice Presidents) in both fast-growing startups and big corporations, and have experimented with different strategies over the years. Oh, the stories we can share…

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 687–688, 2011.

2. How do you conduct product reviews with the executive team?

Come listen to our past experiences to prevent any shock or confusion for the product review attendees.

3. Who should be invited to an executive review meeting?

We will share suggestions to ensure effective meetings with fewer abstract or drifting conversations and more clarifying ones.

4. How do you arrive at decisions during "big stakeholders" reviews?

We'll review the symptoms of a "big stakeholders" meeting gone wrong and share a couple tips on how to make these reviews go more smoothly.

5. How much detail do you include when sharing designs with your executive team?

Learn various tactics to tell compelling stories and help everyone at the table speak the same design language

1.1 Panelists

Selected panelists include a range of roles from product executives to designers in both enterprise and smaller shops.

Jannie Lai, Director of Product Design
Citrix Systems

Chris Maliwat, Vice President, Products
Gilt Groupe

Iram Mirza, Lead Product Designer
Citrix Systems

Nida Zada, Director, Products
Tunerfish

HCI for Peace: Promoting Peace and Preventing War through Computing Technology

Juan Pablo Hourcade[1], Natasha E. Bullock-Rest[1], Janet C. Read[2], and Yoram Chisik[3]

[1] University of Iowa,
Iowa City, IA 52242 USA
[2] University of Central, Lancashire,
Preston, PR1 2HE UK
[3] Universidade da Madeira,
9020-105 Funchal, Portugal
juanpablo-hourcade@uiowa.edu, natasha.bullock.rest@gmail.com,
jcread@uclan.ac.uk, ychisik@gmail.com

Abstract. Our aim in this SIG is to discuss the role human-computer interaction can play in bringing about peace by influencing socio-economic factors that affect the likelihood of conflict as well personal values involved in making decisions to support conflict.

Keywords: Peace, war, pervasive computing, conflict.

1 Introduction

The increasing pervasiveness of computing devices provides human-computer interaction researchers with new opportunities to influence socio-economic and personal factors that affect the likelihood of war and peace. With this in mind, we have started an initiative called HCI for Peace to highlight and celebrate research with peace as its explicit goal. We also hope to form a community of researchers with an interest in using computing technologies to promote peace and prevent war in order to coordinate efforts, share ideas and have a greater impact. Our world can be no brighter than the worlds we dream of.

We have already begun building a community through activities at the CHI conference. During CHI 2010, 500 attendees wore peace ribbons on their name badges and engaged in discussions with others on the role human-computer interaction can play in promoting peace and preventing conflict. During CHI 2011, we are scheduled to present a full paper with a call for constructive action on peace [1], and we will also have a panel discussion. In addition, we have setup a website at hciforpeace.org where we feature interviews and stories on human-computer interaction research for peace.

The INTERACT conference provides us with a unique opportunity to extend our discussions and our community internationally. Through a SIG, we expect to be able to gain new perspectives and bring more voices into the community.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 689–690, 2011.

2 Why Peace?

Peace is a very practical goal to pursue. One reason for this is that war is very expensive. For example, the United States spends roughly 2 billion USD every week on the conflicts in Iraq and Afghanistan [4]. War can also have a devastating economic impact on poorer countries, with most of the poorest countries in the world currently or recently involved in armed conflicts [3]. The most horrible costs though are in terms of human lives, and in the trauma inflicted on those who survive.

3 Discussion

In our previous work [1], we have approached the topic of peace and conflict by reviewing empirical studies on the causes of armed conflict at a socio-economical level. We have also looked into research on moral decisions at a personal level and how these can affect decisions to support war or to kill in the battlefield. Based on this review we have identified human-computer interaction research that is already being conducted and that could be conducted to positively affect both socio-economic factors and personal decisions. Recent events have made these discussions more concrete, with computer technologies, for example, playing a crucial role in (mostly peacefully) toppling dictatorships in Egypt and Tunisia earlier this year [2]. Hosting a SIG will enable us to get new perspectives and discuss alternative ways of approaching research on peace.

A SIG will also be an opportunity to discuss some difficult questions. For example, what role can human-computer interaction realistically play in conflict prevention and resolution? How can we measure the effect of projects on peace and conflict? How can we identify the best situations in which computing technologies can play a positive role? How do we address multiple perspectives during a conflict? We hope these questions will generate fruitful discussions and produce useful ideas.

References

1. Hourcade, J.P., Bullock-Rest, N.E.: HCI for Peace: A Call for Constructive Action. In: Proceedings of CHI 2011, pp. 443–452. ACM Press, New York (2011)
2. Khan, A., Haddad, H., Maher, K., Dubner, R.: Lessons Learned from the Arab Spring Revolutions. In: South by Southwest Conference (2011), http://schedule.sxsw.com/events/event_IAP000461
3. Stewart, F.: Root causes of violent conflict in developing countries. BMJ 324, 342–345 (2002)
4. Stiglitz, J.E., Bilmes, L.J.: The Three Trillion Dollar War: The True Cost of the Iraq Conflict. Norton, New York (2008)

Interaction and Music Technology

Sidney Fels[1] and Michael Lyons[2]

[1] Dept. of Electrical and Computer Engineering, University of British Columbia, 2356 Main Mall, Vancouver, BC, Canada V6T 1Z4
ssfels@ece.ubc.ca
[2] College of Image Arts and Science, Ritsumeikan University, 56-1 Tojiin Kitamachi Kita-ku Kyoto Japan 603-8577
lyons@im.ritsumei.ac.jp

Abstract. This SIG intends to investigate the ongoing dialogue between music technology and the field of human-computer interaction. Our specific aims are to explore major findings of specialized musical interface research over recent years and convey these to HCI researchers who may be interested but not yet active in this area, as well as to consider how to stimulate closer cooperation between music technology and HCI research communities.

Keywords: Music, Technology, Interactive Techniques, HCI.

1 Introduction

Advances in digital audio technologies have led to a situation where computers now play a role in most music production and performance. Digital technologies offer unprecedented opportunities for the creation and manipulation of sound, however the flexibility of these new technologies implies a confusing array of choices for composers and performers. Some musicians have embraced this challenge by using computers directly to create music, leading to an ongoing development of new musical forms. However, most would agree that the computer is not a musical instrument, in the same sense as traditional instruments, and it is natural to seek to design interfaces which allow the computer to be 'played' in fashion suited for human bodies and brains. A decade ago we organized a workshop on "New Interfaces for Musical Expression" [1] at the ACM CHI 2001 conference, with the aim of addressing this issue by exploring connections with the better established field of human-computer interaction. The workshop soon developed into an annual international conference (NIME), which attracts several hundred participants every year. NIME is now considered to be a major venue for the publication of advances in music technology, with an emphasis on methodology and technology from the broader field of human-computer interaction.

2 Relevance to INTERACT 2011

The authors recently began a series of outreach activities [2] to bring a summary of findings of the first ten years of the NIME conference to the attention of the broader

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 691–692, 2011.

research community in human-computer interaction. While some of the members of the music technology research community continue to participate in HCI-related conferences such as CHI and INTERACT, presence of music technology related studies at these conferences is relatively sparse in view of the broad interest in music and its importance for emerging technologies. Moreover we feel that some of the important findings being presented at NIME could be of general interest to researchers who may not have a specific focus on music-related projects.

We therefore propose to invite any participants of INTERACT 2011 who may be interested in music as a potential domain for the study of human-computer interaction to join us for a discussion exploring the past, present, and future dialogue between interaction and music technology.

3 Format of the Discussion

This SIG meeting will consist of a structured discussion addressing several topics of shared interest in the music technology and HCI research communities. Topics for discussion include, but are not expressly limited to the following:

- Design and Aesthetics of musical interface design
- Tools for prototyping new music technology
- Methodologies for evaluating musical interfaces
- Models of 'Playability' of new musical interfaces
- Theories of human action and perception in musical performance
- Mobile music making
- Collaborative music making
- Networked music making
- Music technology for novices
- Musical interfaces in the HCI Curriculum
- Future of Music, Technology, and HCI
- How to better represent music themed research in the HCI community

References

1. Poupyrev, I., Lyons, M.J., Fels, S., Blaine, T.: New interfaces for musical expression. In: CHI 2001 Extended Abstracts on Human Factors in Computing Systems (CHI EA 2001), pp. 491–492. ACM, New York (2001)
2. Fels, S., Lyons, M.: Creating new interfaces for musical expression: introduction to NIME. In: ACM SIGGRAPH 2009 Courses (SIGGRAPH 2009), Article 11, 158 pages. ACM, New York (2009)

User Interface eXtensible Markup Language SIG

Gaëlle Calvary[1], Olivier de Wasseige[2], David Faure[3], and Jean Vanderdonckt[4]

[1] Laboratoire d'Informatique de Grenoble, Equipe Ingénierie de l'IHM
Campus de Grenoble - 385, Rue de la Bibliothèque
BP 53 – F-38041 Grenoble Cedex 9 (France)
gaelle.calvary@imag.fr
[2] Defimedia, Parc Crealys, Rue Phocas Lejeune, 32 – B-5032 Gembloux (Belgium)
olivier.dewasseige@defimedia.be
[3] Thales Research & Technology France - Campus de Polytechnique
F-91767 Palaiseau Cedex (France)
david.faure@thalesgroup.com
[4] Université catholique de Louvain, Louvain School of Management – www.lilab.be
Louvain Interaction Laboratory, Place des Doyens 1 – B-1348 Louvain-la-Neuve (Belgium)
jean.vanderdonckt@uclouvain.be

Abstract. A User Interface Description Language (UIDL) is a formal language used in Human-Computer Interaction (HCI) in order to describe a particular user interface independently of any implementation. Considerable research effort has been devoted to defining various meta-models in order to rigorously define the semantics of such a UIDL. These meta-models cover different aspects: context of use (user, platform, environment), task, domain, abstract user interface, concrete user interface, usability (including accessibility), workflow, organization, evolution, program, transformation, and mapping. A complete development method is typically made up of the following elements: models that capture various aspects of an interactive application (compliant with the previous meta-models), a language that expresses these models, a development life cycle, and software that support this method. This Special Interest Group is aimed at presenting User Interface eXtensible Markup Language (UsiXML), a particular UIDL that is of interest to a wide audience. Then, the UsiXML End User Club is introduced so that any person, group, or organization could observe, test, or contribute to the UsiXML technology. The SIG will present the potential benefits so that everyone can use it.

Keywords: User interface description language (UIDL).

Interest and Relevance

A User Interface Description Language (UIDL) [2,4] is a formal language used in Human-Computer Interaction (HCI) in order to describe a particular User Interface (UI) independently of any implementation technology. As such, a UI may involve different interaction modalities (e.g., graphical, vocal, tactile, haptic, multimodal), interaction techniques (e.g., drag and drop) or interaction styles (e.g., direct manipulation, form filling, virtual reality). A common fundamental assumption of most UIDLs

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 693–695, 2011.

is that UIs are modeled as algebraic or model-theoretic structures that include a collection of sets of interaction objects together with behaviors over those sets. Significant examples of UIDLs include: UIML (www.uiml.org), useML (http:// www.uni-kl.de/pak/useML/), MariaXML, UsiXML (www.usixml.org), and XIML (www.ximl.org). Various UIDLs have been subject to discussion, understanding their common ground and their subsumed approach, a comparative analysis [4], and their consideration for standard [2]. A UIDL can be therefore used during:

- *Requirements analysis*: in order to gather and elicit requirements.
- *Systems analysis*: in order to express specifications those address the aforementioned requirements.
- *System design*: in order to refine specifications depending on the context of use.
- *Run-time*: in order to realize a UI via a rendering engine.

This Special Interest Group (SIG) is manifold:

- To present to a public audience the last version of UsiXML V2.0 (User Interface eXtensible Markup Language – http://www.usixml.org, http://itea.defimedia.be) in terms of coverage (e.g., its support of the Cameleon Reference Framework (CRF) [1] and Similar Adaptation Space (SAS) [5]: who can use UsiXML today and for which purpose?
- To introduce the audience to the models, the language, the syntax, the method, and the software that support the method based on live demonstrations: how to produce a task and domain model that is transformed into an abstract UI (independent of any modality), which is in turn transformed into a concrete UI (independent of any platform, but with a chosen modality) and code generation. This will be illustrated with UI generation in HTML, Java, and Tcl/Tk on multiple platforms (e.g., desktop, laptop, and smartphones).
- To discuss ongoing issues for standardization of user interfaces as defined in the mission statement of the W3C Group on Model-Based User Interface Design []: what are the pros and contras of a standard UIDL
- To invite any individual person, group, community of practice or organization to be part of the **UsiXML End User Club** in order to observe, test, or contribute to the UsiXML technology based on model-driven engineering [6].

References

1. Calvary, G., Coutaz, J., Thevenin, D., Limbourg, Q., Bouillon, L., Vanderdonckt, J.: A Unifying Reference Framework for Multi-Target User Interfaces. Interacting with Computers 15(3), 289–308 (2003)
2. Cantera Fonseca, J.M. (Ed.), Model-based User Interface XG Final Report, W3C Incubator Group Report, W3C (May 4, 2010), accessible at: `http://www.w3.org/2005/Incubator/model-based-ui/XGR-mbui-20100504/`
3. Faure, D., Vanderdonckt, J. (eds.): Proc. of 1st Int. Workshop on User Interface Extensible Markup Language UsiXML 2010 (Berlin) Thales Research and Technology France, Paris (June 20, 2010) ISBN 978-2-9536757-0-2

4. Guerrero-García, J., González-Calleros, J.M., Vanderdonckt, J., Muñoz-Arteaga, J.: A Theoretical Survey of User Interface Description Languages: Preliminary Results. In: Proc. of LA-Web/CLIHC 2009, Merida, pp. 36–43. IEEE Computer Society Press, Los Alamitos (2009)
5. Vanderdonckt, J., Coutaz, J., Calvary, G., Stanciulescu, A.: Multimodality for Plastic User Interfaces: Models, Methods, and Principles. In: Tzovaras, D. (ed.) Multimodal User Interfaces: Signals and Communication Technology. LNEE, pp. 61–84. Springer, Heidelberg (2007)
6. Vanderdonckt, J.: Model-Driven Engineering of User Interfaces: Promises, Successes, and Failures. In: Proc. of 5th Annual Romanian Conference on Human-Computer Interaction ROCHI 2008, Iasi, September 18-19, pp. 1–10. Matrix ROM, Bucharest (2008)

Activity-Centered Interaction Design: A Model-Driven Approach

Larry Constantine

Madeira Interactive Technologies Institute,
University of Madeira, Caminho da Penteada,
9000-390 Funchal, Portugal
LConstantine@m-iti.org

Keywords: activity theory, activity modeling, model-driven design, interaction design.

1 Extended Abstract

Activity theory has had a long history and a growing influence in the design professions broadly conceived and in interaction design in particular. Leading authorities, including Donald Norman, Bonnie Nardi, and others, have called for and argued the advantages of design approaches that focus more on the activities in which human users engage than on the users themselves.

Human activity modeling is a recent development that builds on and enhances the effectiveness of activity theory by enabling more systematic and consistent modeling of core concepts in activity theory. Activity modeling makes possible a model-driven form of activity-centered design. Model-driven approaches are of growing influence in interaction design owing to the promise of yielding more orderly and manageable processes with enhanced traceability from initial conception and the establishment of requirements through to design and final realization.

The tasks that human users perform with software and other designed tools are always performed within the wider context of larger activities that need to be taken into account for the best interaction design. Human activities in the real world are invariably complicated by intricate and largely unpredictable interaction with other systems, artifacts, and other people. Activity theory and human activity modeling provide a simple, powerful framework for thinking about and making sense of how people use tools of all kinds and for setting the stage for delivering better, more usable solutions.

Through human activity modeling, activity-centered design becomes a more disciplined, predictable, and reproducible design process. Using a small set of simplified models and systematized descriptions, human activity modeling provides a concise format in which to capture, carry, and communicate the essence of user needs within human activities.

Human activity modeling provides a bridge between, on the one hand, the interests of interaction designers and other design professionals in understanding users and user needs and designing to support those needs more effectively, and, on the other hand, the interests of software engineers and software developers who most develop

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 696–697, 2011.

the products that actually function to support those needs. Human activity-modeling connects core constructs in activity theory to established concepts and techniques in software engineering, potentially improving communication between designers and developers and leading to a smoother overall development process.

This tutorial will introduce a systematic, activity-centered process for understanding user needs and for designing the interaction between users and software-based systems and services that better support those user needs. This approach grew out of and updates the highly successful and widely used practices of usage-centered design described in the award-winning book, *Software for Use* (Addison-Wesley, 1999). Activity-Centered Interaction Design is model-driven in the sense that simplified models guide and inform the full development process from concept to completed system. The focus is on understanding and modeling the activities in which users engage and then systematically deriving a design that directly and effectively supports users in performing those activities.

Presentations and discussions will be combined with hands-on small-group exercises and a case study problem will enable participants to experience and trace the entire model-driven design process through to a preliminary design.

Topics to be covered in this tutorial include:

- understanding human activity:
  from activity theory to activity modeling
- human activity modeling:
  context models, participation models, and performance models
- understanding and modeling users within activities:
  user roles, personas, and role profiles
- activity-based task modeling:
  from activities to essential-form use cases
- model-based user interface architecture:
  navigation and content organization
- model-driven abstract prototyping:
  beyond wireframe schematics
- principle-driven interface design:
  model-driven derivation of design details

References

1. Constantine, L.L.: Human Activity Modeling: Toward a Pragmatic Integration of Activity Theory and Usage-Centered Design. In: Seffah, A., Vanderdonckt, J., Desmarais, M. (eds.) Human-Centered Software Engineering II. Springer, NY (2008)
2. Constantine, L.L.: Beyond User-Centered Design and User Experience. Cutter IT Journal 17(2) (February 2004)
3. Constantine, L.L., Windl, H.: Usage-Centered Design: Scalability and Integration with Software Engineering. In: Stephanidis, C., Jacko, J. (eds.) Proceedings of the 10th International Conference on Human-Computer Interaction: Theory and Practice, Crete, Greece. Lawrence Erlbaum Associates, Mahwah (2003)
4. Constantine, L.L., Lockwood, L.A.D.: Usage-Centered Engineering for Web Applications. IEEE Software 19(2), 42–50 (2002)

Analysis, Redesign and Evaluation with Teasing Apart, Piecing Together

Clare J. Hooper

Eindhoven University of Technology, P.O. Box 513,
5600 MB, Eindhoven, Netherlands
c.j.hooper@tue.nl

Abstract. This half-day tutorial will teach participants how and when to use Teasing Apart, Piecing Together (TAPT), a two-phase design method for understanding and redesigning user experiences in new contexts. TAPT was developed to address a gap in the field for methods oriented around experiences, particularly with respect to understanding their social and emotional facets. TAPT has been successfully used in both industry and academia, and this workshop will draw on the tutor's experiences in the field.

Keywords: TAPT, UX, understanding, analysis, design, evaluation.

1 Teasing Apart, Piecing Together

TAPT concerns understanding and redesigning experiences in new contexts. It has been used to facilitate the design of real-world versions of experiences that were initially situated on the web (such as microblogging and wiki usage). TAPT's development was motivated by issues of accessibility, such as the lack of access to web-based social tools by people who are offline.

TAPT falls into two phases:

1. 'Teasing Apart' involves analysing an experience's 'surface elements' (design / physical aspects) and 'experienced effects' (literal and abstract outcomes). These are reviewed to identify essential aspects and produce a 'distilled' description that does not refer to the original modality.
2. In 'Piecing Together' practitioners take the distilled experience description and use it as a springboard for creatively redesigning the initial experience in the new context, which may be physical or digital.

2 The Tutorial

Overview: The first half of the tutorial covers the basics of TAPT, demonstrating how it can be used to solve design problems. This session will draw on information from a 43-participant trial of TAPT conducted across industry and academia.

The second half of the tutorial concerns agile uses of the method. Participants will learn how TAPT has been used in the field for analysis and evaluation as well as for

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 698–699, 2011.

design, and about different modes of usage: with end users or practitioners, and by groups or individuals. This section of the day will draw on knowledge gained from fieldwork, specifically case studies of TAPT's successful usage in academia (in the UK and Norway) and industry (by IBM UK and IBM India).

This tutorial combines formal presentations with group exercises and feedback. Participants will be asked to work in groups to respond to design tasks, and to briefly present their results to one another. Participants will be encouraged to provide feedback to one another (within and between groups), and to engage with the tutor.

Learning outcomes:

- how to analyse design problems with TAPT
- how to use TAPT for redesign
- how to evaluate TAPT-built designs
- how to design TAPT workshops, considering aspects such as available resources, number and background of participants, and workshop format

Audience: Design practitioners, HCI researchers and practitioners, students, other professionals in the field.

Prior knowledge required: None.

Supporting materials: participants will be be given copies of tutorial materials: slides introducing TAPT and explaining modes of use; TAPT instructions and forms; sample TAPT exercises. Participants will also receive copies of key publications[1].

Tutor: Clare J. Hooper is a computer scientist based at the Eindhoven University of Technology. During her doctorate, she developed and evaluated TAPT. In addition to design, Clare's research interests include web science, hypertext and HCI.

3 Provisional Schedule

15 minutes	Opening words and settling in
45 minutes	Presentation / Q&A. Why use TAPT; what it looks like; the types of problem it can solve
30 minutes	Group work: a simple redesign exercise
15 minutes	Feedback and discussion of group work
15 minutes	Coffee break
30 minutes	Presentation / Q&A. Agile uses of TAPT: analysis, design and evaluation; use with practitioners or end users; creativity prompts
30 minutes	Group work: choice between analysis or evaluation exercises
15 minutes	Feedback and discussion of group work
15 minutes	Final questions, closing remarks

[1] Hooper, C. J. and Millard, D. E. (2010) Teasing Apart and Piecing Together: Towards Understanding Web-based Interactions. *Web Science 2010*, April 2010, Raleigh, USA. Hooper, C. J. and Rettberg, J. W. (2011) Using TAPT as an Analytical Method for Understanding Online Experiences. *Web Science 2011*, June 2011, Koblenz, Germany.

Context-Aware Adaptation of User Interfaces

Vivian Genaro Motti and Jean Vanderdonckt

Université catholique de Louvain, Louvain School of Management
Louvain Interaction Laboratory, Place des Doyens 1
B-1348 Louvain-la-Neuve, Belgium
{vivian.genaromotti,jean.vanderdonckt}@uclouvain.be

Abstract. Efficient adaptation aims at ensuring that a user interface is adapted to a user's task according to the context of use, since the end user is carrying out a task with one or several computing platforms in a physical environment. This tutorial presents key concepts of adaptation: principles that guide it, relevant context information and how to consider it, dimensions and abstraction levels subject to adaptation, as well as, languages, methods and techniques used in this domain. This tutorial aims at teaching major aspects to be considered for adaptation of user interfaces in general and concerning the context of use in particular, including the end user (or several of them, as in multi-user interfaces), the platform (or several of them, as in multi-device environments), and the physical environment (or several of them, as in multi-location systems).

Keywords: Multi-dimensional adaptation, Context-aware adaptation, Adaptive applications, Adaptable applications.

1 Tutorial Information

Adaptation changes system aspects according to its context and defined rules. Many dimensions can be subject to it (as navigation and presentation), and in different granularity levels, ranging from atomic units to the whole application. Efficient adaptation provides users a better interaction by considering the context. It may be performed by rules and supported by methods, techniques, languages and machine learning approaches. Methods combine adaptation techniques, e.g. to adapt the content, a text can be summarized, simplified and described. Languages that respect separation of concerns (user interface description languages) are recommended. The application domains for adaptation vary: entertainment, education, e-commerce. Contributions in this field have been reported since the early 90's, yet, the evolution of the technologies requires constant efforts to provide adaptation efficiently. Also, current users interact with new devices (e.g. idTV), via different means (e.g. touchscreens), using new technologies (e.g. RIAs). The varied contexts, dimensions and levels for adaptation are a challenge. Users may feel confused or overloaded with excesses. Today, a large body of knowledge about adaptation exists but it is very scattered and widespread in the literature, requiring time and efforts to find relevant information when one wishes to support some form of user interface adaptation (adaptability and/or adaptivity). This tutorial addresses this need providing an overview of user interface adaptation in general and of context-aware adaptation in

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 700–701, 2011.

particular. It covers the state-of-the-art of multi-dimensional context-aware adaptation, in scientific and commercial domains, and it is organized in:

Fundamental Concepts. The diversity of contexts of use is a *challenge* for the development of systems that suits to user' preferences, profiles, and requirements. Designing different versions of a system for each scenario is a possible *strategy* to perform adaptation, though it demands effort and may lead to inconsistent results. *Models* of different abstraction levels can also be created to accommodate varied scenarios [1]. Adaptation aims at a universal access, independent of user, platform, and environment. The *context of use* is the source of information for adaptation, it can be *automatically* or *manually* gathered. Many *dimensions* of a system can be subject to adaptation, and in different *granularity levels*. The context defines the priority of each dimension. *Principles*, as plasticity, *rules* and *techniques* support this process.

Theoretic Models. *Taxonomies* and *ontologies* have been used to classify, define and model context, and adaptation techniques. The *context-aware design space* (CADS) organizes dimensions and levels for adaptation. Orthogonal axis defines granularity levels for each dimension. Adaptive/adaptable applications can be compared with CADSs, which are descriptive, extensible and flexible. A *reference framework* organizes context information for adaptation, defining where, when, how adaptation can be performed. To adapt system resources many *techniques* can be applied, a *template* details them. Adaptation *models* describe *views* of a system, and the recommended languages in this sense, consider *separation of concerns*. *Animation* is one possible *strategy* to present the adaptation result smoothly for users [2].

Use Cases. In commercial domain examples of application include: (i) the Google search engine that adapts the order of the results according user preferences; Google AdSense that provides context-aware advertisements; (ii) Amazon recommends products according to the shops of other costumers; (iii) NYTimes suggests articles according to their content; and (iv) Nokia Situations that allows users to define context settings for device acts. In scientific domain, examples include: (i) The Sedan-Bouillon website [3] that adapts its content according to the platforms and user preferences, (ii) QTk Draw [4] that allows users to distribute the application in different devices, and (iii) ELM-ART [5] that adapts the navigation for the user.

References

1. Motti, V.G.: A Computational Framework for Multi-dimensional Context-aware Adaptation. In: Proc. of EICS 2011, Pisa, June 13-16. ACM Press, New York (2011)
2. Dessart, C.-E., Motti, V., Vanderdonckt, J.: Showing User Interface Adaptivity by Animated Transitions. In: Proc. EICS 2011, Pisa. ACM Press, New York (2011)
3. Balme, L., Demeure, A., Calvary, G., Coutaz, J.: Sedan-Bouillon: A Plastic Web Site. In: PSMD 2005, INTERACT 2005 Workshop on Plastic Services for Mobile Devices (2005)
4. Grolaux, D., Vanderdonckt, J., Van Roy, P.: Attach me, detach me, assemble me like you work. In: Costabile, M.F., Paternó, F. (eds.) INTERACT 2005. LNCS, vol. 3585, pp. 198–212. Springer, Heidelberg (2005)
5. Brusilovsky, P., Schwarz, E.: ELM-ART: An Intelligent Tutoring System on the WWW

Designing the Search Experience

Tony Russell-Rose

UXLabs Ltd. and City University London
Northampton Square, London EC1V 0HB
tgr@uxlabs.co.uk

Abstract. This half-day tutorial provides a practical introduction to Human-Centred Design for information search, access and discovery. We present a concise overview of the fundamental concepts and principles of human information-seeking behaviour and show how to apply these in the design of search user experiences. A key element of the tutorial is the opportunity to practice these skills in a group exercise.

Keywords: search, navigation, information retrieval, information discovery, data visualization, user experience, user-centred design.

1 Tutorial Goals

Participants in this half-day tutorial will:

- Explore the fundamental concepts and principles of Human-Centred Design for information search, access and discovery;
- Learn how to differentiate between various types of search behaviour: known-item, exploratory, etc.;
- Study models of human information-seeking behaviour (e.g. Broder, Norman, Marchionini, Bates, etc.), and how to apply interaction design principles based on those models;
- Develop an understanding of the key dimensions of user type, goal and mode of interaction, and how to apply these variables when designing for different user contexts;
- Understand the role of design patterns, and how to apply UI design patterns from various libraries in designing search user interfaces;
- Gain an awareness of the key design resources available within the HCIR community and how to apply these to practical design challenges.

2 Content

1. **Introductions and objectives:** Group introductions & Ice-breaker. A brief summary of what each participant hopes to gain from the session, and what experiences they bring.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 702–703, 2011.

2. **Understanding Search & Discovery Behaviour:** An overview of the key models and frameworks for human-information seeking, focusing on the work of Marchionini, Bates, and more contemporary works of Morville, Tunkelang at al.
3. **Faceted Classification & Search:** A review of Ranganathan's seminal work on Colon Classification and an exploration the implications for the design of contemporary faceted classification and search paradigms.
4. **Varied Solutions for Varied Contexts:** An exploration of the universal dimensions that define search and discovery experiences, and how these translate into principles for the design of contemporary faceted search experiences.
5. **UI Design Pattern Libraries:** A detailed examination of best practices in search experience design, embodied as design patterns in key HCIR design pattern libraries.
6. **Exercise: UX Review:** An opportunity to practice all the above skills in a group exercise. The scenario will involve critical review of a live site, with analysis and review of the search and discovery experience, and synthesis of appropriate design insight and recommendations.
7. **Exercise: UX Review (Feedback & Presentations):** Feedback and review of the group exercise.
8. **Conclusions & Wrap-up:** A review of the overall session, including the shared experiences of the group exercises and the contrasting findings of each.

3 Intended Audience

This tutorial is aimed at information architects, user experience architects, search specialists, and HCI practitioners and researchers interested in the designing more effective user experiences for search and information discovery.

4 Presenter's Background

Tony Russell-Rose is currently director of UXLabs, a consultancy specialising in user experience research, design and analytics. Before founding UXLabs he was Manager of User Experience at Endeca and editor of the Endeca UI Design Pattern Library, an online resource dedicated to best practice in the design of search and discovery experiences.

Prior to this he was technical lead at Reuters, specialising in advanced user interfaces for information access and search. And before Reuters he was R&D group manager at Canon Research Centre Europe, where he led a team developing next generation information access products and services. Earlier professional experience includes a Royal Academy of Engineering fellowship at HP Labs and a Short-term Research Fellowship at BT Labs. His academic qualifications include a PhD in human-computer interaction, an MSc in cognitive psychology and a first degree in engineering, majoring in human factors. Tony is also Honorary Visiting Fellow at the Centre for Interactive Systems Research, City University London.

Improving the Content of User Requirements

Nigel Bevan

Professional Usability Services, 12 King Edwards Gardens, London W3 9RG, UK
mail@nigelbevan.com

Abstract. Identifying and defining user requirements is an essential input to good user centred design, but there is little guidance on content. The workshop will share and review examples of user requirements provided by the participants, to generate a contents list that could help practitioners identify and document the relevant requirements.

Keywords: Usability, requirements.

1 Description

Most existing approaches emphasize the need to understand user requirements in order to provide a basis for good design. They describe the methods that can be used to gather information about users and their tasks (e.g. [1,2]). For example the methods suggested by Courage and Baxter [2] are interviews, surveys, user needs analysis, card sorting, group task analysis, focus groups and field studies. But what are the resulting requirements against which the completed system could be evaluated? The Common Industry Specification for Usability Requirements [3] emphasizes the value of high-level requirements for effectiveness, efficiency and satisfaction. But how should these be complemented by more detailed requirements?

An ISO standards group is attempting to define the contents of a user requirements specification, but this has proved to be unexpectedly difficult to do.

The objective of the workshop is to identify a set of categories of user requirements that cover a range of projects and organizations, which could be used to help practitioners elicit, identify and document the relevant requirements. Participants should have experience of producing user requirements, and provide an (anonymized) example at least one month before the workshop, together with their own suggestions for categorization. Issues to be discussed include:

- Is a common categorization possible or appropriate across different organizations and application domains?
- Is there a difference between user needs and user requirements, and is the same categorization appropriate for both?
- How important is it to include requirements for effectiveness, efficiency and satisfaction?
- Should user requirements always be documented, and is it possible in principle to evaluate whether they have been achieved?

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 704–705, 2011.

The most appropriate method for wider dissemination will be discussed at the end of the workshop, with a paper in Interactions or the Journal of Usability Studies being possibilities. The results will also submitted to the ISO standards group for information, and could influence national comments on the draft standards.

1.1 Organizer

Nigel Bevan is an independent consultant and researcher with wide industrial experience. He has contributed to many international standards, and is a member of the ISO standards group that is developing a Common Industry Format for User Requirements Specifications. Nigel is co-editor of the Common Industry Format for Context of Use Descriptions and the new version of the ISO standard for usability methods supporting human-centered design. He was a member of the National Academy of Science Committee on Human-System Design Support for Changing Technology.

2 Schedule

Overview of the examples and summary of prior discussion and conclusions from review of the examples before the workshop.
Brief presentation of each example with questions for clarification.
Discussion: what are the common principles and situation-specific factors?
Identify topics for detailed discussion, ideally in groups, for example based on application domain, methodology or organization size.
LUNCH
Discussion of topics.
Report back/suggestions for a contents list that categorizes needs and requirements (guided by the issues identified). The conclusions may be general, or specifically related to particular situations.
Plans for dissemination and any further work.

The timetable provides a framework for discussing these issues, but will be adapted depending on the knowledge and experience of the participants, and the issues that arise from the position papers.

References

1. Bevan, N., Maguire, M.: User requirements analysis: A review of supporting methods. In: Proceedings of IFIP 17th World Computer Congress (2002)
2. Courage, C., Baxter, K.: Understanding your users. Elsevier, Amsterdam (2005)
3. NIST: Common Industry Specification for Usability—Requirements (2007), `http://nist.gov/iusr`

Model-Driven Inquiry: Beyond Ethnography and Contextual Inquiry

Larry Constantine

Madeira Interactive Technologies Institute,
University of Madeira, Caminho da Penteada,
9000-390 Funchal, Portugal
LConstantine@m-iti.org

Keywords: model-driven inquiry, user research, user requirements, contextual inquiry, ethnography.

1 Extended Abstract

Model-driven approaches are of growing influence in interaction design owing to the promise of yielding more orderly and manageable processes with enhanced traceability from initial conception and the establishment of requirements through to design and final realization. Model-driven inquiry is an agile technique, an accelerated alternative in its own right to contextual inquiry and other ethnographic approaches for user research, field study, and requirements gathering that can also be combined with these more conventional techniques.

Model-driven inquiry is a proven, industrial-strength method that can streamline and systematize user research. The approach enables projects to get the greatest benefit from the fewest resources by reducing the amount of time needed for adequate field investigation. The approach allows for focusing field inquiry more sharply and specifically by identifying specific areas for investigation based on risk, uncertainty, ambiguity, and impact on design outcomes. By drawing on the extant knowledge and insight of designers and identifying boundaries between the known and not-known, model-driven inquiry can dramatically reduce the need for extended observation or extensive user research. Although the approach evolved within interaction design practice, it also has potential application in academic research.

Participants in this tutorial will learn about the relationship of model-driven inquiry to contextual inquiry and other methods based in ethnography and participant observation. Through a mix of presentation, discussion, hands-on application to real-world problems, and review of experiences, participants will learn how to use models and modeling to drive inquiry, how to choose models for leverage in both inquiry and design, and how to formulate focused and efficient field research by selecting effective inquiry techniques. The relationship to and integration with contextual inquiry will also be addressed.

Model-driven inquiry, based on exploratory modeling and using techniques from joint essential modeling, has been used extensively in design practice and is taught in the joint Carnegie Mellon/University of Madeira MHCI program.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 706–707, 2011.

Presentations and discussions will be combined with hands-on small-group exercises and a case study problem will enable participants to experience and trace the entire model-driven design process through to a preliminary design.

Topics to be covered in this tutorial include:

- nature and philosophy of ethnography and ethnographic methods
- overview of contextual inquiry
- model-driven inquiry, exploratory modeling, and joint essential modeling
- exploratory models: inventories, maps, and profiles
- activity and task modeling, basic definitions
- methods of generating inventory content:
- activities, roles, and tasks
- provisional drafts of activity profiles and role profiles
- user and professional participation in exploratory modeling
- compiling, clustering, and categorizing queries
- confidence sorting, salience sorting, and risk ranking of queries
- selecting appropriate inquiry methods and target informants
- incorporating inquiry findings to refine and elaborate models
- iterative inquiry for agile projects
- integrating model-driven inquiry with contextual inquiry

References

1. Constantine, L.L.: Model-Driven Inquiry. User Experience 8(3), 16–19 (2009)
2. Constantine, L.L.: Human Activity Modeling: Toward a Pragmatic Integration of Activity Theory and Usage-Centered Design. In: Seffah, A., Vanderdonckt, J., Desmarais, M. (eds.) Human-Centered Software Engineering II. Springer, NY (2008)
3. Constantine, L.L.: Beyond User-Centered Design and User Experience. Cutter IT Journal, 17(2) (February 2004)
4. Windl, H.: Usage-Centered Exploration: Speeding the Initial Process. In: Constantine, L. (ed.) Proceedings, for USE 2002: First International Conference on Usage-Centered, Task-Centered, and Performance-Centered Design, pp. 39–53. Ampersanda Press, Rowley (2002)

Scenario-Based Requirements Engineering Facilitating Interaction Design

Hermann Kaindl

Vienna University of Technology,
Gußhausstr. 27-29, A-1040 Vienna,
Austria
kaindl@ict.tuwien.ac.at

Abstract. When the requirements and the interaction design of a system are separated, they will most likely not fit together, and the resulting system will be less than optimal. Even if all the real needs are covered in the requirements and also implemented, errors may be induced by human-computer interaction through a bad interaction design and its resulting user interface. Such a system may even not be used at all. Alternatively, a great user interface of a system with features that are not required will not be very useful as well.

Therefore, the primary motivation of this tutorial is to improve system development in practice both regarding requirements engineering and interaction design, especially facilitating the latter. We argue for combined requirements engineering and interaction design, primarily based on usage scenarios in the sense of sequences of actions aimed at accomplishing some task goal. However, scenario-based approaches vary especially with regard to their use, e.g., employing abstract use cases or integrating scenarios with functions and goals in a systematic design process. So, the key issue to be addressed is how to combine different approaches, e.g., in scenario-based development, so that the result is an overall useful and useable system. In particular, scenarios are very helpful for purposes of usability as well.

Keywords: Interaction design, usage scenarios, requirements engineering, user interfaces, usability.

1 Tutorial Goals

This tutorial is targeted towards people who are supposed to work on the interaction design or the requirements in systems development, e.g., interaction designers, user interface developers, Web designers, requirements engineers, or project managers. Whatever the roles of the tutorial participants actually are in their daily work, they should get a better understanding of "other" viewpoints and tasks and, in particular, a common approach. The overall goal of presenting this proposed tutorial is to teach how requirements engineering and interaction design relate and how they can be usefully combined. This can be important for creating better interactive systems in the future.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 708–709, 2011.

2 Key Learning Outcomes

In this tutorial, participants learn about an approach to scenario-based requirements engineering that facilitates interaction design. In particular, participants will understand how scenarios and use cases can be utilized both for requirements engineering and interaction design, though with different emphasis on the level of detail. They will also understand the additional need to specify the functional requirements for the system to be built, even in the context of object-oriented (OO) development. Overall, they will gain a better understanding of early systems design.

3 CV of the Presenter

Hermann Kaindl is the director of the Institute of Computer Technology at the Vienna Univ. of Technology. He joined this institute in early 2003 as a full professor. Prior to moving to academia, he was a senior consultant with the division of program and systems engineering at Siemens AG Austria. There he has gained more than 24 years of industrial experience in software development and human-computer interaction. He has published four books and more than a hundred papers in refereed journals, books and conference proceedings. He is a *Senior Member* of the IEEE, a *Distinguished Scientist* member of the ACM, a member of the INCOSE and the AAAI, and is on the executive board of the Austrian Society for Artificial Intelligence.

References

1. Kaindl, H.: A Practical Approach to Combining Requirements Definition and Object-Oriented Analysis. Annals of Software Engineering 3, 319–343 (1997)
2. Kaindl, H.: A Design Process Based on a Model Combining Scenarios with Goals and Functions. IEEE Transactions on Systems, Man, and Cybernetics (SMC) Part A 30, 537–551 (2000)
3. Kaindl, H.: Adoption of Requirements Engineering: Conditions for Success. In: Proceedings of the Fifth IEEE International Symposium on Requirements Engineering (RE 2001), invited State-of the-Practice Talk, Toronto, Canada, pp. 156–163. IEEE, Los Alamitos (2001)
4. Kaindl, H.: Is Object-oriented Requirements Engineering of Interest? Requirements Engineering 10, 81–84 (2005)
5. Kaindl, H.: A Scenario-Based Approach for Requirements Engineering: Experience in a Telecommunication Software Development Project. Systems Engineering 8, 197–210 (2005)
6. Kaindl, H., Jezek, R.: From Usage Scenarios to User Interface Elements in a Few Steps. In: Proceedings of the Fourth International Conference on Computer-Aided Design of User Interfaces (CADUI 2002), Valenciennes, France, pp. 91–102. Kluwer Academic Publishers, Dordrecht (2002)
7. Kaindl, H., Kramer, S., Hailing, M.: An Interactive Guide Through a Defined Modelling Process. In: People and Computers XV, Joint Proceedings of HCI 2001 and IHM 2001, Lille, France, pp. 107–124. Springer, London (2001)
8. Kaindl, H., Svetinovic, D.: On confusion between requirements and their representations. Requirements Engineering 15, 307–311 (2010)

Sketching Interactive Systems with Sketchify

Željko Obrenović

User Centered Engineering Group, Department of Industrial Design,
Eindhoven Technical University (TU/e)
Tel.: +31 40 247 5245
z.obrenovic@tue.nl
http://www.idemployee.id.tue.nl/z.obrenovic/

Abstract. Sketching is at the heart of design and creativity, an omnipresent element of any disciplined activity of design. In this tutorial we will summarize many of the existing studies of sketching, and emphasize its role in supporting creativity. We will look at how sketching aids in reflection and conversation, and supports designers' memory and cognition. We will discuss the relation of sketching to prototyping and engineering, and present Sketchify, a software tools for sketching beyond paper and pencil. We will cover various techniques that can be used to extend sketching to other forms than simple creation of a pencil trace on paper.

Keywords: Sketching, design of interactive systems, creativity.

1 Motivation

Sketching is at the heart of design and creativity. Many studies of design practice, such as more recent contributions from Buxton [Buxton 2007], Krippendorff [Krippendorff 2006], and Moggridge [Moggridge 2007], have called attention to sketching as an omnipresent element of any disciplined activity of design.

In this tutorial we will summarize many of the existing studies of sketching, and emphasize its role in supporting creativity. We will look at how sketching aids in reflection and conversation, and supports the designers' memory and cognition. We will discuss the relation of sketching to prototyping and engineering, and present some software tools for sketching.

We will address the need for extending sketching to other forms than simple plain paper. This need is especially evident in the domain of interaction design, where designers also need means to deal with attributes of the overall user experience, such as time, phrasing, and feel. While disciplines such as graphical design and architecture have a rich practice of sketching and courses that students can take in order to improve their sketching skills, interaction designers cannot efficiently employ existing sketching techniques while designing new classes of user interfaces. We will present some novel ideas and approaches that that try to go beyond these limitations and redefine sketching as a creativity support tool that can be supported through various media.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 710–711, 2011.

2 Topic Covered

Part 1: Sketching, Creativity and Interaction Design

- What is Sketching?
- Sketching as a Medium for Reflection and Conversation
- Cognitive Psychology Viewpoint: Sketching as an Extension of Designer Memory and Cognition
- Towards Abstract Attributes of Sketching
- Sketching vs. Prototyping and Engineering

Part 2: Sketchify: Techniques for Sketching Beyond Paper and Pencil

- Visual effects: animation (flipbook, properties), page transitions (screen prototyping)
- Simplified programming: events, timers, macros, scripts
- Working with I/O services: Web cameras, sensors, Wii remote, Web services, Arduino...
- Hacking techniques: screen capturing, screen poking, screen scraping
- Connecting to external devices: iPhone example

References

1. Obrenović, Z., Martens, J.B.: Sketching Interactive Systems with Sketchify. ACM Trans. on CHI 18(1) (March 2011)
2. Buxton, B.: Sketching User Experiences: Getting the Design Right and the Right Design. Morgan Kaufmann, San Francisco (2007)
3. Moggridge, B.: Designing interactions. MIT Press, Cambridge (2007)
4. Krippendorff, K.: The semantic turn: A new foundation for design. Taylor & Francis, Boca Raton (2006)

UIs Automatically Optimized for Your Smartphone

Hermann Kaindl

Vienna University of Technology
Gußhausstr. 27-29, A-1040 Vienna, Austria
kaindl@ict.tuwien.ac.at

Abstract. Graphical user interfaces (UIs) for PCs will most likely not fit relatively small screens of devices like today's Smartphones. Providing dedicated UIs for several devices manually, however, is costly and takes time. Therefore, we have developed an approach to (semi-)automatic generation of UIs for various devices. A designer defines classes of dialogues in a device-independent discourse model. Such a discourse model can be also viewed as specifying classes of scenarios, i.e., use cases. It refers to a domain model that specifies the domain-of-discourse of the dialogues between user and application. From such models, we can generate UIs (semi-)automatically. Recently, we included in this generation process automatic optimization based on heuristic search. In effect, this tutorial shows that and how user interfaces can be automatically optimized for your Smartphone.

Keywords: (Semi-)automatic generation of user interfaces, automatic optimization for small devices, Smartphones.

1 Tutorial Goals

The main goals are to show that usable *user interfaces* (UIs) can be generated (semi-)automatically, in particular through *automatic optimization* based on heuristic search. This approach is particularly useful for optimizing UIs for relatively small screens like those of current *Smartphones*. Just based on a simple device specification of the Smartphone, the UI is specifically optimized for it. Such UIs can be generated fully automatically and optimally fit the given device, so that they are ready for real-world use, e.g., on Smartphones.

2 Key Learning Outcomes

In this tutorial, participants learn about modeling discourses using a new approach inspired by human-human communication. They will know how modeling discourses and generating user interfaces can be approached systematically. They will also see how techniques from heuristic search can be used for optimizing automatically generated user interfaces for small devices like today's Smartphones.

3 CV of the Presenter

Hermann Kaindl is the director of the Institute of Computer Technology at the Vienna Univ. of Technology. He joined this institute in early 2003 as a full professor. Prior to

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 712–713, 2011.

moving to academia, he was a senior consultant with the division of program and systems engineering at Siemens AG Austria. There he has gained more than 24 years of industrial experience in software development and human-computer interaction. He has published four books and more than a hundred papers in refereed journals, books and conference proceedings. He is a *Senior Member* of the IEEE, a *Distinguished Scientist* member of the ACM, a member of the INCOSE and the AAAI, and is on the executive board of the Austrian Society for Artificial Intelligence.

References

1. Bogdan, C., Falb, J., Kaindl, H., Kavaldjian, S., Popp, R., Horacek, H., Arnautovic, E., Szep, A.: Generating an Abstract User Interface from a Discourse Model Inspired by Human Communication. In: Proceedings of the 41st Annual Hawaii International Conference on System Sciences (HICSS-41). IEEE Computer Society Press, Big Island (2007)
2. Bogdan, C., Kaindl, H., Falb, J., Popp, R.: Modeling of interaction design by end users through discourse modeling. In: Proceedings of the 2008 ACM International Conference on Intelligent User Interfaces (IUI 2008). ACM Press, Maspalomas (2008)
3. Falb, J., Kaindl, H., Horacek, H., Bogdan, C., Popp, R., Arnautovic, E.: A discourse model for interaction design based on theories of human communication. In: CHI 2006 Extended Abstracts on Human Factors in Computing Systems, pp. 754–759. ACM Press, New York (2006)
4. Falb, J., Kavaldjian, S., Popp, R., Raneburger, D., Arnautovic, E., Kaindl, H.: Fully Automatic User Interface Generation from Discourse Models. In: Proceedings of the 2009 ACM International Conference on Intelligent User Interfaces (IUI 2009). ACM Press, Sanibel Island (2009); Tool demo paper
5. Falb, J., Popp, R., Röck, T., Jelinek, H., Arnautovic, E., Kaindl, H.: UI Prototyping for Multiple Devices Through Specifying Interaction Design. In: Baranauskas, C., Abascal, J., Barbosa, S.D.J. (eds.) INTERACT 2007. LNCS, vol. 4662, pp. 136–149. Springer, Heidelberg (2007)
6. Kaindl, H.: A Design Process Based on a Model Combining Scenarios with Goals and Functions. IEEE Transactions on Systems, Man, and Cybernetics (SMC) Part A 30, 537–551 (2000)
7. Kaindl, H., Jezek, R.: From Usage Scenarios to User Interface Elements in a Few Steps. In: Proceedings of the Fourth International Conference on Computer-Aided Design of User Interfaces (CADUI 2002), Valenciennes, France, pp. 91–102. Kluwer Academic Publishers, Dordrecht (2002)
8. Kavaldjian, S., Bogdan, C., Falb, J., Kaindl, H.: Transforming Discourse Models to Structural User Interface Models. In: Giese, H. (ed.) MODELS 2008. LNCS, vol. 5002, pp. 77–88. Springer, Heidelberg (2008)
9. Raneburger, D., Popp, R., Kavaldjian, S., Kaindl, H., Falb, J., Optimized, G.U.I.: Optimized GUI Generation for Small Screens. In: Model-Driven Development of Advanced User Interfaces, MDDAUI 2010 Workshop. Springer, Heidelberg (2011)

User Experience Evaluation – Which Method to Choose?

Virpi Roto[1], Arnold Vermeeren[2], Kaisa Väänänen-Vainio-Mattila[3], and Effie Law[4]

[1] University of Helsinki, P.O.Box 68, 00014 University of Helsinki, Finland
virpi.roto@helsinki.fi
[2] Delft University of Technology, Landbergstraat 15, 2628 CE Delft, The Netherlands
a.p.o.s.vermeeren@tudelft.nl
[3] Tampere University of Technology, Korkeakoulunkatu 6, 33720 Tampere, Finland
kaisa.vaananen-vainiomattila@tut.fi
[4] University of Leicester, LE1 7RH Leicester, UK
elaw@mcs.le.ac.uk

Abstract. User experience has many dimensions and therefore, it is tricky to evaluate it. When the goal of user experience evaluation is to investigate how people feel about using an interactive system, the traditional usability methods are hardly applicable. In this tutorial, we introduce a set of 78 user experience evaluation methods that we have been collecting from the user experience community 2008-2010. We give both an overview of the different types of methods and examine a selected set of methods in detail.

Keywords: User experience, Evaluation, Assessment, Method.

1 Introduction

High quality user experience (UX) has become a central competitive factor of product development in mature consumer markets. Improving UX of systems requires evaluation, but traditional usability testing methods are not adequate for evaluating UX. The evaluation methods for investigating how users *feel* about the tested system are still largely unknown.

Since 2008, we have been collecting a comprehensive set of UX evaluation methods (UXEM) both from academia and industry. We have collected the methods from workshops [2,3,4], a Special Interest Group session [1], online survey, literature, and existing smaller collections of experiential evaluation methods.

Based on a structured description of each method, we have been able to categorize the methods based on various criteria. Figures 1 and 2 show examples of UXEMs categorized by the time span of UX that is investigated and by the product development phase they can be used in.

In this tutorial, we will share our knowledge on UX evaluation methods by both giving an overview of the different types of UXEMs and by examining a selected set of methods to gain practical understanding of the method.

By the end of this tutorial, participants will know

- the general targets of UX evaluation
- the various kinds of UX evaluation methods available for different purposes
- how to choose the right method for the purpose

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 714–715, 2011.

Evaluating emotions	Evaluating an episode	Evaluating long-term UX
Observation	**Observation**	**Self-reporting**
Facial, body, vocal expressions (e.g. smile, lean back, sigh)	Experience think aloud	Questionnaires, Laddering, iScale, Repertory Grid Technique
Psychophysiological measurements	**Self-Reporting**	
Muscle, pupil, heart, skin reactions detected with sensors	Experience sampling, AttrakDiff, Interviews, Day Reconstruction	
Self-reporting		
Verbal: PANAS, AffectGrid Non-verbal: EmotionSlider, EmoCards, PrEmo		

Fig. 1. Examples of UX evaluation methods for studying different periods of experience

Concepting	Non-functional prototype	Functional prototype
Visual design	**Visual design**	**Lab test**
Emotional expressions,reactions	Emotional expressions,reactions	Emotional expressions,reactions AttrakDiff
Idea description	**Interaction**	**Field study**
Expert evaluation, Role play: Perspective-Based Inspection	Experience think aloud	Experience sampling, Diary, Day Reconstruction Method
		Market feedback
		Questionnaires, UX Curve / iScale

Fig. 2. Examples of UX evaluation methods for different phases of product development

- the basics of the selected UX methods of different types
- where to find more information on those methods

References

1. Obrist, M., Roto, V., Väänänen-Vainio-Mattila, K.: User Experience Evaluation - Do you know which method to use? In: Extended Abstracts of CHI 2009, pp. 2763–2766 (2009)
2. Roto, V., Hoonhout, J.: Experiential Evaluation Methods. In: Workshop in the 4th International Conference on Designing Pleasurable Products and Interactions, DPPI 2009 (2009)
3. Roto, V., Väänänen-Vainio-Mattila, K., Law, E., Vermeeren, A.: User experience evaluation methods in product development (UXEM'09). In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5727, pp. 981–982. Springer, Heidelberg (2009)
4. Väänänen-Vainio-Mattila, K., Roto, V., Hassenzahl, M.: Now Let's Do It in Practice: User Experience Evaluation Methods in Product Development. In: CHI 2008 Proceedings - Workshops, Florence, Italy, April 6, pp. 3961–3964 (2008), Workshop site `http://www.cs.tut.fi/ihte/CHI08_workshop`

User Experience Evaluation in Entertainment and Games

Regina Bernhaupt

IRIT, Groupe ICS; 118 Route de Narbonne,
31062 Toulouse, France
Regina.Bernhaupt@irit.fr

In a nutshell. This tutorial comprehensively covers important user experience (UX) evaluation methods and opportunities and challenges of UX evaluation in the area of entertainment and games. The course is an ideal forum for attendees to gain insight into state-of-the art user experience evaluation methods, going way-beyond standard usability and user experience evaluation approaches in area of human-computer interaction. It surveys and assesses the efforts of user experience evaluation of the gaming and human computer interaction communities during the last 10 years.

Keywords: entertainment, user experience, evaluation methods, beyond usability, game.

1 Introduction

User experience evaluation in games and more general in interactive entertainment system has become a focus of (research) attention in the areas of human-computer interaction (HCI). During the last 10 years the scientific communities of HCI and game research were starting to learn from each other: On the one hand UX evaluation methods from HCI are used during the game development, on the other side HCI was borrowing and investigating aspects of the gaming experience like immersion, flow or fun, to better understand the concept of UX [1].

This course is about evaluating user experience in the area of entertainment and (video) games. It provides an overview on what user experience is about (contrary to usability), it provides an understanding on what makes for successful future games and entertainment experiences and what set of user experience evaluation methods is currently available and used for the development of games.

Objective of this course is to provide *an overview* on user experience evaluation in the games and entertainment area. The goal is to provide some *definitions* of user experience, explain the *factors* that might be contribution to the overall user experience in a game (e.g. *flow, immersion, playability*) and (using a set of examples) to explain how games development is slightly different from standard software engineering development. Based on these foundations the objective is to give an overview on existing methods from the field and allow participants in the course a first hands-on experience trying out how to apply one of the methods to a real game.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 716–717, 2011.

2 Schedule and Presentation Format

The tutorial will use both presenter-led material and group-based exercises throughout the sessions. Small exercises and hands-on examples will encourage attendees both to critique existing approaches as well as posing their own view on the relative importance of the various factors that are contributing to user experience. The course will include demonstrations of UX evaluation methods and use video clips to show the variety of applications domains the methods can be applied to. Course materials will include copies of the presentation materials as well as extensive follow-up reading lists.

The tutorial will run over a half day. The material will focus on an overview on available evaluation methods, their application in the game-development life cycle and shortcomings and benefits for each of the methods. The course consists of four blocks:

1. Introduction to the topic, definition of user experience in entertainment and games, application areas
2. Overview on user experience evaluation methods
3. Application of user experience evaluation (hands-on exercise) and wrap up of insights from the exercise
4. Key insights summary and lessons learned

3 Target Audience

This course will appeal to a broad audience. For *developers and designers*: the course will help understand how to evaluate user experience in the area of games and entertainment and how outcomes of the evaluation can be integrated in the next iteration of the game and entertainment application development; for *industrial and academic researchers*: The course will provide an overview on current methods in that area, and can help understand the concept of user experience; for *students*: the course provides a first introduction to user experience in games, but lessons can be taken up for the application in other domains (e.g. definition of UX, application of methods during a development life cycle). The course additionally can be of interest to any attendee of Interact who wants to get a first idea and understanding on user experience and user experience evaluation in general.

References

1. Bernhaupt, R.: User Experience Evaluation in Games: Concepts and Methods. Springer, Heidelberg (2010)

5th Workshop on Software and Usability Engineering Cross-Pollination: Patterns, Usability and User Experience

Peter Forbrig[1], Regina Bernhaupt[2], Marco Winckler[3], and Janet Wesson[4]

[1] University of Rostock, Computer Science Department, Albert-Einstein-Str. 21. D-18051 Rostock, Germany
Peter.Forbrig@uni-rostock.de
[2] ruwido, user experience research, Köstendorferstr. 17 A-5202 Neumarkt, Austria
Regina.Bernhaupt@ruwido.com
[3] IRIT, Groupe ICS, 118 Route de Narbonne 31062 Toulouse, France
Marco.Winckler@irit.fr
[4] Nelson Mandela Metropolitan University, Dept. of Comp. Sc. & Inf. Syst., PO Box 77000 6031 Port Elisabeth, South Africa
Janet.Wesson@nmmu.ac.za

Abstract. The workshop focuses on how process models, methods and knowledge from the area of Human-Computer Interaction can be integrated and adopted to support and enhance traditional software engineering processes. In its 5th edition this workshop will investigate the application of usability engineering methods that are adapted to fit the evaluation of advanced interfaces and how usability and user experience evaluation methods can be incorporated to support design decisions and changes in standard software development. This workshop is organized by the IFIP working group 13.2 "Methodologies for User-Centered Systems Design".

Keywords: Software Engineering, Usability, User Experience, Cross-pollination, Patterns.

1 Introduction

Software engineering and usability engineering are affected by a mutual influence that we call “cross-pollination”. Examples are task specifications, design patterns and life cycle models. These examples were invented in one field and later on adapted in a new context. Use cases [5] and Usability pattern languages [7] are only two out of many examples. New developments in intelligent and adaptive environments and mobile computing require new solutions, for usability evaluation methods [3] and especially for user experience evaluation [1]. The key attribute of user interfaces is that they need to adapt to time, location and usage which makes them very difficult to evaluate using standard techniques [2].

The workshop will focus on how to integrate and extend traditional development and evaluation methods in order get user interfaces that are usable and ensure good user experience. Additionally, it should be possible to optimally evaluate the usability

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 718–719, 2011.

of advanced interfaces in their specific context of use [4, 6]. Experts in HCI, software and usability engineering need to learn from each other to facilitate and encourage this convergence.

The workshop aims to be a forum for sharing ideas about potential and innovative ways to cross-pollinate the expertise among the different communities and to show examples, which can stimulate industrial software development. Additionally it should provide a forum that will help to grow a community of interest in this area.

2 Structure of the Workshop

The goals of this workshop are to provide HCI specialists, software engineers and usability specialists from industry and research institutions the opportunity to discuss both the state-of-the art and the cutting edge practice in usability and user experience evaluation. Topics of interest include, but are not limited to, the usability and user experience evaluation of advanced interfaces and interactive systems like adaptive interfaces, context-aware interfaces, human-robot interfaces or mobile interfaces and the integration of these methods in the respective application domains. Additionally, reports about the application of patterns in the different fields of HCI like [7] are welcomed.

The workshop is the official workshop of IFIP working group 13.2 "Methodologies for User-Centered Systems Design". http://wwwswt.informatik.uni-rostock.de/IFIP_13_2/. It expects HCI specialists, software and usability engineers from academia and industry as participants.

This workshop is planned for one full day including the following activities of an invited talk, papers presentations and a round table discussion. Participants have to prepare a position paper of 4 to 10 pages which will be reviewed by an international committee. Selected papers will be published on the workshop web site (http://CEUR-WS.org.) and will be presented during the workshop. The outcome of the workshop will be a white paper presented on the web site of the workshop.

References

1. Bernhaupt, R. (ed.): Evaluating User Experience in Games: Concepts and Methods. Springer, London (2010)
2. Brusilovsky, P., Karagiannidis, C., Sampson, D.: The Benefits of Layered Evaluation of Adaptive Applications and Services. In: Bauer, M., Gmytrasiewicz, P.J., Vassileva, J. (eds.) UM 2001. LNCS (LNAI), vol. 2109. Springer, Heidelberg (2001)
3. Chin, D.N.: Empirical Evaluation of User Models and User-Adapted Systems. User Modelling and User Adapted Interaction (11), 181–194 (2001)
4. Gena, C.: Methods and Techniques for the Evaluation of User-Adaptive Systems. The Knowledge Engineering Review 20(1), 1–37 (2005)
5. Jacobson, I.: Object- Oriented Software Engineering. In: A Use Case Driven Approach. Addison Wesley, Reading (1992)
6. Kjeldskov, J., Graham, C.: A Review of Mobile HCI Research Methods. In: Chittaro, L. (ed.) Mobile HCI 2003. LNCS, vol. 2795. Springer, Heidelberg (2003)
7. Winkler, M., Bernhaupt, R., Pontico, F.: Challenges for the development of user interface pattern languages: A case study on the e-Government domain. IADIS International Journal on WWW/Internet 8(2), 59–84 (2010)

Accessible Design in the Digital World

Gerhard Weber[1], Helen Petrie[2], and JennyDarzentas[3]

[1] Department of Computer Science, Technical University of Dresden
Nöthnitzer Straße 46, 01187 Dresden, Germany
Gerhard.Weber@tu-dresden.de
[2] Department of Computer Science, University of York,
Deramore Lane, Heslington East,
York YO10 5GH United Kingdom
Helen.Petrie@cs.york.ac.uk
[3] Department of Product and Systems Design Engineering, University of the Aegean,
Konstantinoupoleos, 2 Hermoupolis, Syros Greece
jennyd@aegean.gr

Abstract. The workshop provides an opportunity for researchers, practitioners and designers interested in eAccessibilityto discuss and debate the possibilities for accessibility and usability in the emerging world of Web 2.0, ubiquitous and pervasive technologies, and multimodal interaction, bridging desktop and mobile applications.

1 Background to the Workshop

Many approaches for developing user interfaces are based on the concept of user centered design (UCD). If UCD is applicable, the capabilities of users, analysis of tasks and contexts are the basis for all further work including the evaluation of some interactive system. An approach for designing accessible user interfaces is much harder to implement as in all steps users with large variations of capabilities and preferences have to be considered.

Communities of users have unique ways of addressing issues and different approaches for solving problems arising from lack of accessibility. Web 2.0 addresses communities of users specifically and allows authoring of content in flexible ways. But authoring of Web 2.0 content by users with different disabilities is still difficult and new approaches to implement assistive technologies will be needed to ensure successful integration of these users.

2 Goals and Issues

The workshop aims to develop a better understanding of the varied communities of users in the digital world, how accessible design solutions benefit from collaboration of users having possibly different if not contradicting requirements, systems that support such collaboration and evaluation methods allowing better insight by experts into the limitations of assistive technologies while using services in the digital world collaboratively.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 720–721, 2011.

Web 2.0 has arrived with a plethora of possibilities to contribute information through many types of media. The workshop is interested in studies accessibility issues related to usage of such services.

Mobile devices combine social platforms with location based services. Thereby the digital and the physical world may be bridged while taking the physical context of services and applications into account. In this context, lack of mobility adds to the possible of accessibility of mobile services. Reporting about obstacles, collaborative pre-journey systems, finding accessible routes and rating places from the perspective of users with a disability or who are elderly are topics of this workshop.

Novel cyber-physical systems (Lee, 2008) integrate computational with physical design, propose more natural user interfaces, may be multimodal, and are implemented as embedded systems. Accessibility of cyber-physical devices is much harder to ensure as embedded devices lack assistive technologies by and large. Tangible user interfaces are one example. Access by physically impaired people to collaborative tangible interaction for example on large multitouch tables cannot easily be ensured by writing software and must be designed from the beginning on. Submissions to this workshop should address solutions for designing accessible cyber-physical systems.

Support tools and methods in all stages of user-centered design are to be considered in Web 2.0 application domains such as education, location-based services, and e-government. Cloud-based computing and collaborative accessibility are examples for new technologies which may be utilized for better integration of all users.

3 Organisation

The Accessible Design in the Digital World conference was conceived by the University of Dundee's Digital Media Access Group, which was held in Dundee in 2005. ADDW 2008 was hosted by York University and brought an international group together, who were presenting their papers over two days (Petrie, 2008).

ADDW 2011 is supported by IFIP Working Group TC 13.3 *HCI and Disabilities*. Organizers of ADDW 2008 and chairs of WG 13.3 are organizing this workshop.

References

Lee, E.: Cyber Physical Systems: Design Challenges. University of California, Berkeley Technical Report No. UCB/EECS-2008-8 (January 23, 2008), `http://www.eecs.berkeley.edu/Pubs/TechRpts/2008/EECS-2008-8.html`

Petrie, H. (ed.): Proc. ADDW 2008, CD-ROM, University of York, Dept. Computer Science (2008)

Building Bridges – HCI and Visualization

Achim Ebert[1], Gitta Domik[2], Nahum Gershon[3], and Gerrit van der Veer[4]

[1] University of Kaiserslautern, Gottlieb-Daimler-Strasse, 67663 Kaiserslautern, Germany
ebert@cs.uni-kl.de
[2] University of Paderborn, Fürstenallee 11, 33102 Germany
domik@uni-paderborn.de
[3] The MITRE Corp., 7515 Colshire Drive, McLean, VA 22102-7508, USA
gershon@mitre.org
[4] Open University the Netherlands, Valkenburgerweg 177, 6419 AT Heerlen, The Netherlands
gerrit@acm.org

Abstract. The fields, HCI and visualization, are usually practiced as two separate disciplines by researchers with different backgrounds and capabilities. However, these two disciplines, HCI and visualization, could complement each other and leveraging on the differences and complementary features of the two research fields could be beneficial for both. In this workshop, we are going to discuss the different approaches and capabilities of these two disciplines and layout a road map for a unified approach of research using both.

Keywords: HCI, Visualization, Standardization.

1 Workshop Objectives

Whenever discussing the relation between HCI and visualization in general or when presenting research results in these areas, questions arise about the differences between these research fields. Aren't both fields just the same? And if not, where is the common ground? Where is one research field dependent on the other? Can we combine the separated viewpoints and paradigms in a unified and complementary approach, or are we forced to choose one or the other? How can we provide the general public (the developers and users of visualization and HCI and the engineers implementing our designs) a precise and practical enough idea about what's happening in these fields and what's not? What are the consequences of the answers on the previous: how and what should we teach? What will be the future? This dilemma is a topic of frequent discussion around the water cooler, lecture halls, as well as in the boardroom.

One of the major issues is that it is not easy to precisely define the terms visualization and HCI and that there are many interpretations of these two fields that appear to be distinct.

In order to better (or at all) answer the questions on similarities, differences, and correlations of HCI and Visualization, in our workshop we want to discuss topics like:

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 722–723, 2011.

- What is HCI? What is Visualization? What is a working description that is practical highlighting the special features of each of the fields?
- Are there other disciplines involved in this struggle (e.g., Visual Analytics)?
- How can we take advantage of the two fields and how can we find ways for people with different inclinations to collaborate and take advantage of the strengths of each other?
- What are the similarities of the disciplines? What are the major differences?
- Do we need to really split the domains? Or do we need to join them and provide a joint curriculum for studying and practicing them?
- Can we give definitions that are better applicable in real situations?
- Does one need to further research the ways to make people take advantage of both disciplines in designing interactive visual systems? In that case, what are the research agenda(s) and what are the Top 10 Research Challenges?

We welcome participants from various backgrounds interested in research and application of HCI and visualization, including designers, artists, researchers in visualization, interaction, psychology, and usability, and people from all application fields. The intended length of the workshop is one day. Prospective attendees will submit open position papers from their own areas of interest and also provide short answers to two pro forma questions asking for (i) the participant's views on the most important existing knowledge in the area, including a position statement on possible definitions, and (ii) key research challenges related to HCI and Visualization issues.

The team of organizers is comprised of representatives of both university and industry, giving them a wide multi-disciplinary expertise. They all have significant experience in the main disciplines, as well as in related areas and application domains. Some of the organizers have already worked together in many workshops of the HCIV series [*www.hciv.de*]. The gained expertise and large number of members will be of a great value for successfully advertising this workshop.

Beside the position papers submitted by the workshop attendances, we plan to start an interactive blog in order to continue the discussions of the workshop. For a better visibility of our actions and progresses made, we will also distribute and discuss them using means of social media like Facebook groups and Twitter. Furthermore, the results of the discussions should form common ground for at least one high-quality conference or journal paper.

2 Additional Information

Additional information (full proposal, first schedule, references, CfP) on this proposed workshop can be found at *http://www.hciv.de/interact11*

Combining Design and Engineering of Interactive Systems through Models and Tools (ComDeisMoto)

Stefan Sauer[1], Kai Breiner[2], Heinrich Hussmann[3], Gerrit Meixner[4], Andreas Pleuss[5], and Jan Van den Bergh[6]

[1] University of Paderborn, s-lab – Software Quality Lab, D-33095 Paderborn, Germany
[2] University of Kaiserslautern, Erwin-Schroedinger-Str, D-67663 Kaiserslautern, Germany
[3] University of Munich, Amalienstr. 17, D-80333 Munich, Germany
[4] DFKI, Trippstadter Str. 122, D-67663 Kaiserslautern, Germany
[5] Lero, University of Limerick, Limerick, Ireland
[6] UHasselt – IBBT, Wetenschapspark 2, 3590 Diepenbeek, Belgium
sauer@s-lab.upb.de, breiner@cs.uni-kl.de, heinrich.hussmann@ifi.lmu.de, gerrit.meixner@dfki.de, andreas.pleuss@lero.ie, jan.vandenbergh@uhasselt.be

Abstract. Development of interactive systems and their user interfaces combines engineering and design, formal and informal development methods from different domains. Diverse models and tools can be used to support the developers' work. In model-driven development, software systems are specified systematically using dedicated models for different aspects of the system. Yet, appropriate design of user interfaces is as important as functional correctness. This workshop provides a forum of multi-disciplinary discussion on how to integrate model-driven development with the often more informal methodologies used in user-centered design and engineering. Starting point of the discussion are the tools, models, methods and experiences of the workshop participants.

Keywords: User-interface engineering, interaction design, model-driven development, user-centered design, models.

1 Model-Driven Development of Interactive Systems

Model-Driven Development (*MDD*) has become an important paradigm in software development. In MDD, models are the primary artifacts in the development process. Models are often visual models, like UML models, but can also be represented in textual formats like XML. Each model is denoted by a modeling language that can be tailored to the according stakeholders while still remaining machine applicable.

Analogous to traditional software development, MDD processes leverage models on different levels of abstraction. But unlike traditional software development, the transition between the levels can be described in a formal way and enable (semi-) automatic transformations between the levels, even finally into implementation source code. Having models as development artifacts also enables the application of model checking methods for verification, early testing or tool support.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 724–725, 2011.

MDD provides a large number of powerful concepts and tools to deal with models, meta-models, and model transformations. Together, they provide support for systematic and efficient software development.

Model-driven development of interactive systems (MDDIS) and, in particular, their user interfaces (UI) applies the principles of MDD to the target domain of interactive systems and UIs. While traditional m*odel-based user interface development* approaches have not spread widely into practice in the past, requirements of new interaction devices, e.g. demanding for device independence, make model-driven approaches more and more important. One major advantage of MDD is that there are different models for particular purposes. One well-known use case is to separate the content (*what* is displayed) from the design (*how* it is displayed) into distinct models.

This workshop follows the series of workshops on model-driven development of advanced user interfaces (MDDAUI), held in 2005-2010. In the five editions of that workshop series we have seen a lot of models, transformations and tools for model-driven user interface development. These approaches also tackled advanced UI features like multi-platform development and plasticity, context-sensitiveness, multimedia, 3D and augmented reality, ambient production environments, wearable sensors, interactive TV, and many others.

2 Design and Engineering: Enhancing the User Experience

The previous workshop editions have shown that concepts for model-driven user interface development are already becoming mature. However, the quality of the resulting UIs, in terms of the user experience, has always been one of the most important challenges. In some cases, the UIs generated from the models can even improve the usability as they are for instance very consistent. However, it comes to a drawback if some parts of the UI require individual design e.g. because of its complexity or to increase the likeability of the product.

For these reasons, an optimal development method should support both: systematic MDDIS and incorporation of individual design knowledge, the latter usually resulting from manual, informal methods in user-centered design.

3 Workshop Goals

This workshop focuses on challenges, opportunities, practical problems, and proposed solutions to integrate MDDIS and informal design methods and tools. It aims at fostering interdisciplinary discussions between the respective expert groups and viewpoints. Workshop participants investigate how models, tools and current or future MDD techniques can be combined with design expertise in a way that improves the development process for interactive systems, better supports the work of UI creators and eventually benefits the user experience of end users. Our objective is improving the tools for software engineers and UI developers, being the creators and users of models, e.g. by integrating design measures in engineering models, providing modeling tools for designers or novel methods for MDD, etc.

Data-Centric Interactions on the Web

Paloma Díaz[1], Tim Hussein[2], Steffen Lohmann[1], and Jürgen Ziegler[2]

[1] Universidad Carlos III de Madrid,
Computer Science and Engineering Department,
Avenida de la Universidad 30, 28911 Leganés, Spain
{pdp,slohmann}@inf.uc3m.es
[2] University of Duisburg-Essen,
Department of Computer Science and Applied Cognitive Science,
Lotharstrasse 65, 47048 Duisburg, Germany
{tim.hussein,juergen.ziegler}@uni-due.de

Abstract. The World Wide Web has become a global database in recent years, with an ever-growing amount of data that is published online every day. Interactions with this data raise a number of research questions and practical challenges that are not yet sufficiently investigated. These issues will be addressed by the International Workshop of Data-Centric Interactions on the Web. It will serve as a platform for researchers, developers, and designers to foster the exchange of experiences and to discuss novel research ideas and results regarding data-centric interactions on the web.

Keywords: Web interaction, user interfaces, data-centric systems, semantic web, visualization, large datasets, interactive systems, data management.

1 Workshop Theme

While the original World Wide Web was mainly a web of documents, today's web is characterized by an ever-growing amount of data that is published and linked in structured formats. Data about products or services is increasingly made available for public access by companies (e.g., Amazon A2S API, Google Maps API, etc.). Governments and organizations publish more and more often statistics and other data on the web, usually with the aim to increase transparency and empowering the public to utilize this data (e.g., data.gov, data.gov.uk, etc.). In other attempts, structured data is extracted from Wikipedia (e.g., DBpedia [1, 2]) or from multiple web sources (e.g., Freebase [3, 4]). Last but not least, huge amounts of data and metadata are created by the web users themselves, either indirectly (e.g., via social tagging) or in directed community efforts (e.g., Open Directory Project, OpenStreetMap, etc.). These are just a few of the many examples where data is published on the web nowadays.

Several standards and best practices for the description, publication, linking, and exchange of web data have been developed in the past. Popular examples are W3C specifications such as XML, SOAP, and RDF, the recommended best practices of the Linked Data initiative [5], and various vocabularies that emerged from these approaches (e.g., Dublin Core, SKOS, etc.). A similar line is taken in less formal attempts to structure data, such as microformats [6] or advanced tagging approaches (e.g., geotagging, hashtags, etc.).

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 726–727, 2011.

However, the full potential of the data can only be exploited with well-designed user interfaces and powerful interaction techniques that allow an efficient exploration and utilization of the data. Although some early attempts to investigate the interaction with this data have been made in the past, there is still a large number of research questions and practical challenges that are not yet sufficiently addressed. Reoccurring interaction problems are differently solved and more general design recommendations and guidelines are just beginning to emerge. A need for reusable design patterns and interaction techniques as well as novel ideas, tools, and methods to present web data to the users is clearly recognizable. These and related issues of data-centric interactions on the web are addressed by the workshop.

2 Format and Target Audience

The workshop will be a full-day event consisting of paper presentations, discussions, tool demonstrations, and collaborative activities. It aims to bring together researchers, developers, and designers with a shared interest in data-centric interactions on the web. As several research areas are affected by the workshop theme, participants from various backgrounds and communities are expected to attend. The workshop will provide an interdisciplinary forum to exchange experiences and ideas and to discuss novel research results and future developments regarding data-centric interactions on the web.

The workshop will consist of 3-4 sessions, each focusing on a particular aspect of the workshop theme. Accepted papers will be presented in 15-20 minute time slots. A summarization of all papers at the end of each session will help to identify shared goals and challenges and stimulate an in-depth discussion of the session topics. One session will be dedicated to tool demonstrations and selected design cases. In a final wrap up session, open problems and a roadmap for future research efforts will be identified. Further information can be found on the workshop website at http://dci-workshop.org

References

1. Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R., Ives, Z.G.: DBpedia: A Nucleus for a Web of Open Data. In: Aberer, K., Choi, K.-S., Noy, N., Allemang, D., Lee, K.-I., Nixon, L.J.B., Golbeck, J., Mika, P., Maynard, D., Mizoguchi, R., Schreiber, G., Cudré-Mauroux, P. (eds.) ASWC 2007 and ISWC 2007. LNCS, vol. 4825, pp. 722–735. Springer, Heidelberg (2007)
2. DBpedia, http://dbpedia.org
3. Bollacker, K., Evans, C., Paritosh, P., Sturge, T., Taylor, J.: Freebase: a Collaboratively Created Graph Database for Structuring Human Knowledge. In: Proc. of SIGMOD 2008, pp. 1247–1250. ACM, New York (2008)
4. Freebase, http://www.freebase.com
5. Linked Data – Connect Distributed Data across the Web, http://linkeddata.org
6. Microformats, http://microformats.org/

Encouraging Serendipity in Interactive Systems

Stephann Makri[1], Elaine G. Toms[2], Lori McCay-Peet[2], and Ann Blandford[1]

[1] UCL Interaction Centre, 8th Floor, MPEB, University College London, Gower Street, London WC1E 6BT, UK
{S.Makri,A.Blandford}@ucl.ac.uk
[2] iLab, Dalhousie University, 6100 University Avenue, Halifax, NS, B3H 3J5, Canada
{etoms,mccay}@dal.ca

Abstract. We regularly make serendipitous discoveries in both online and offline contexts – from stumbling upon a useful website when searching for something completely different to meeting someone with mutual research or business interests in an unlikely place. However, most existing interactive systems do not provide a fertile environment for serendipity to occur. This workshop will identify key requirements and research challenges for designing and evaluating user-centred systems that aim to encourage serendipity.

1 Introduction and Workshop Objectives

While not usually life-changing like Fleming's discovery of Penicillin, we often make serendipitous discoveries that involve a mix of chance and insight and lead to a valuable outcome – , particularly when interacting with information and people. For example, consider finding a useful article for a paper you are currently writing while researching a different topic entirely, meeting someone with mutual business or research interests in an unlikely place such as at a conference on a different subject, or applying for a job position after being approached by a company that notices your enthusiastic tweets on Twitter.

Interactive systems cannot directly 'induce' serendipity. Indeed, it can be argued that even 'designing to *encourage*' it is an oxymoron; as soon as serendipity is 'engineered' into a system, discoveries might cease to be regarded as serendipitous even if they lead to a valuable outcome. This is because the system is likely to reduce the amount of chance and insight involved in the serendipitous discovery simply by helping to encourage it. So how can we harness the potential of this fleeting concept? How can we best design interactive systems that surprise and delight users without devaluing users' perceptions of the role of chance or insight? And how can we best evaluate the success of the systems we design if by designing to encourage serendipity, we might actually make the resultant discoveries appear to be *less* serendipitous than if they had been made without our assistance? With these tensions in mind, this workshop will identify key requirements and research challenges for designing and evaluating user-centred systems that aim to encourage serendipity.

The workshop will be structured according to 3 themes: 1) ***understanding*** serendipity to inform design (i.e., what do we know and need to know about the conditions that stimulate serendipity?), 2) ***designing*** to encourage serendipity (i.e., what are the key requirements for designing interactive systems so that conditions are

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 728–729, 2011.

optimum for serendipity to occur?), and 3) ***evaluating*** the success of interactive systems aimed at encouraging serendipity (i.e., what approaches and criteria should we use to judge the success of the systems we design?).

2 Workshop Structure and Organisation

The workshop will begin with 1 minute introductions by each participant, focusing on serendipity-related interests. Next, multi-disciplinary groups of 3-4 chosen by the organisers based on attendees' research interests will sub-divide by the 3 workshop themes to discuss the *current state-of-the-art* for understanding serendipity, designing to encourage it and evaluating the success of interactive systems that aim to encourage it. Findings will be reported back to the workshop and the links across themes will be discussed as a whole group. After lunch, a group design activity will allow participants to consider how serendipity might be enabled by *future* technologies. Individuals will be asked to think of a novel interactive tool or system aimed at encouraging serendipity, then pair up with someone from a different institution with a similar/complementary idea. The process will be repeated to form groups of 4 and these groups asked to devise a scenario of how their tool/system might be used and reflect on the key requirements and research challenges highlighted by their scenario for understanding serendipity and designing and evaluating systems aimed at encouraging it. Each group will present their scenario and a summary of the requirements and research challenges identified. This will be followed by a plenary discussion on how the identified requirements and challenges might be addressed through future research efforts. Finally, there will be an optional group dinner.

3 Target Audience and Expected Workshop Outcomes

To encourage a range of perspectives during discussion, and fitting with this year's conference theme, this workshop aims to 'build bridges' by bringing together up to 20 participants from a mix of HCI and related disciplines (e.g. Design, Psychology, Information Retrieval, Information Science and Computer Science) who share an interest in the *user-focused* design and evaluation of interactive systems that aim to encourage serendipity. We expect most participants to be researchers or PhD students as well as practitioners who are currently involved in developing interactive systems.

We have created a workshop website to facilitate discussion, knowledge-sharing and post-workshop collaboration (http://sites.google.com/site/serendipityworkshop). The current approaches and research challenges identified for each workshop theme and future research directions will be summarised on the site and will form the basis of an article to be submitted to *ACM Communications* and one aimed at communicating our findings to practitioners in a magazine such as *Interactions*.

Acknowledgements. This work is partly supported by EPSRC Project EP/H042741/1 to Blandford, and NCE GRAND and a SSHRC grant to Toms.

Human Work Interaction Design for e-Government and Public Information Systems

Dinesh Katre[1], Pedro Campos[2], Torkil Clemmensen[3],
Rikke Orngreen[4], and Annelise Mark Pejtersen[5]

[1] Human-Centred Design & Computing, C-DAC Pune, India
[2] Madeira Interactive Technologies Inst., Universidade da Madeira, Portugal
[3] Department of Informatics, Copenhagen Business School, Denmark
[4] Media and ICT in a Learning Perspective, Aarhus University, Denmark
[5] Center of Cognitive Systems Engineering, Denmark
dinesh@cdac.in, pedro.campos.pt@gmail.com, tc.inf@cbs.dk,
RIOR@dpu.dk, ampcse@mail.dk

Abstract. Varied backgrounds of users, heterogeneous delivery media and diverse socio-cultural and organizational contexts pose new challenges of human work interaction design in the field of e-government and public information systems. The workshop will consolidate the empirical case studies of how human work analysis and interaction design has benefited in enhancing the user experience of e-government and public information systems and a set of effective methods, techniques and theories for this purpose. Selected research papers from the workshop will be published in the International Journal of Public Information Systems (IJPIS), Sweden.

Keywords: Interaction design, human work analysis, empirical case studies, e-government, public information systems.

1 Workshop Theme

This proposal is for the **one day** workshop during INTERACT 2011 to be organized by the members of WG 13.6 Human Work Interaction Design (authors of this proposal). We propose to involve the community of WG 8.5 Information Systems in Public Administration.

The complex interplay and information exchanges between the models of Government to Government, Government to Business and Government to Citizen pose different kind of challenges while designing the e-government systems. Human work analysis is very critical in public information systems design as the procedures, working styles, organizational contexts are different in various public sector organizations. Interaction design of e-government system and public information systems has to simultaneously address the user experience requirements of diverse users such as people from varied educational and professional backgrounds, senior citizens, physically disabled, illiterate and people with different languages and cultural preferences. The combination of Human Work analysis and Interaction Design in user interface design for e-government systems has to adopt itself to diverse delivery media such as web, desktop, touch screen kiosks and mobile devices.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 730–731, 2011.

We would like to invite position papers on the theme of the workshop following (but not excluded to) methodological approaches and domains such as –

- Human work analysis and interaction design case studies in e-government and public information systems
- The influence of socio-economical, political and cultural context on the design of public information systems
- Social networks and collaborative user interfaces for policy crowd-sourcing
- Government-centric versus citizen-centric design approaches
- User interface design for diverse delivery media for e-government services such as web, desktop, touch screen kiosk and mobile devices
- Designing for diverse users, Every-citizen interface design, Mass personalization
- Democratic design for e-government
- HCI challenges in bridging the digital divide through e-government
- Usability evaluation methods for e-government
- User experience design for e-government

1.1 Intended Audience

The theme of this workshop involves many disciplines across academic boundaries and hence a diverse audience can submit papers and attend the workshop. Obviously, people attending INTERACT 2011 and Participants of IFIP HWID WG 13.6, but also practitioners and researchers, who attend only this workshop, because they have a specific interest in the proposed themes. For example, researchers and practitioners in:

- e-government and public information systems
- Organizational engineering and work analysis
- Interaction design
- E-government policy makers
- Public information system design / development
- HCI, Usability and UX practice

This INTERACT workshop takes place in Portugal, and we encourage Portuguese researchers and practitioners to submit papers and to participate.

1.2 Publication of Workshop Papers

We have already formed an international programme committee for reviewing the research papers of this workshop. International Journal of Public Information Systems (IJPIS), Sweden has agreed to publish a special issue from the proceedings of this workshop.

Improving the Content of User Requirements

Nigel Bevan

Professional Usability Services, 12 King Edwards Gardens, London W3 9RG, UK
mail@nigelbevan.com

Abstract. Identifying and defining user requirements is an essential input to good user centred design, but there is little guidance on content. The workshop will share and review examples of user requirements provided by the participants, to generate a contents list that could help practitioners identify and document the relevant requirements.

Keywords: Usability, requirements.

1 Description

Most existing approaches emphasize the need to understand user requirements in order to provide a basis for good design. They describe the methods that can be used to gather information about users and their tasks (e.g. [1,2]). For example the methods suggested by Courage and Baxter [2] are interviews, surveys, user needs analysis, card sorting, group task analysis, focus groups and field studies. But what are the resulting requirements against which the completed system could be evaluated? The Common Industry Specification for Usability Requirements [3] emphasizes the value of high-level requirements for effectiveness, efficiency and satisfaction. But how should these be complemented by more detailed requirements?

An ISO standards group is attempting to define the contents of a user requirements specification, but this has proved to be unexpectedly difficult to do.

The objective of the workshop is to identify a set of categories of user requirements that cover a range of projects and organizations, which could be used to help practitioners elicit, identify and document the relevant requirements. Participants should have experience of producing user requirements, and provide an (anonymized) example at least one month before the workshop, together with their own suggestions for categorization. Issues to be discussed include:

- Is a common categorization possible or appropriate across different organizations and application domains?
- Is there a difference between user needs and user requirements, and is the same categorization appropriate for both?
- How important is it to include requirements for effectiveness, efficiency and satisfaction?
- Should user requirements always be documented, and is it possible in principle to evaluate whether they have been achieved?

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 732–733, 2011.

The most appropriate method for wider dissemination will be discussed at the end of the workshop, with a paper in Interactions or the Journal of Usability Studies being possibilities. The results will also submitted to the ISO standards group for information, and could influence national comments on the draft standards.

1.1 Organizer

Nigel Bevan is an independent consultant and researcher with wide industrial experience. He has contributed to many international standards, and is a member of the ISO standards group that is developing a Common Industry Format for User Requirements Specifications. Nigel is co-editor of the Common Industry Format for Context of Use Descriptions and the new version of the ISO standard for usability methods supporting human-centered design. He was a member of the National Academy of Science Committee on Human-System Design Support for Changing Technology.

2 Schedule

Overview of the examples and summary of prior discussion and conclusions from review of the examples before the workshop.
Brief presentation of each example with questions for clarification.
Discussion: what are the common principles and situation-specific factors?
Identify topics for detailed discussion, ideally in groups, for example based on application domain, methodology or organization size.
LUNCH
Discussion of topics.
Report back/suggestions for a contents list that categorizes needs and requirements (guided by the issues identified). The conclusions may be general, or specifically related to particular situations.
Plans for dissemination and any further work.

The timetable provides a framework for discussing these issues, but will be adapted depending on the knowledge and experience of the participants, and the issues that arise from the position papers.

References

1. Bevan, N., Maguire, M.: User requirements analysis: A review of supporting methods. In: Proceedings of IFIP 17th World Computer Congress (2002)
2. Courage, C., Baxter, K.: Understanding your users. Elsevier, Amsterdam (2005)
3. NIST: Common Industry Specification for Usability—Requirements (2007), `http://nist.gov/iusr`

Mobile Accessibility Workshop

Daniel Gonçalves[1], Luis Carriço[2], and Markel Vigo[3]

[1] Instituto Superior Técnico / TULisbon,
Rua Alves Redol 9, 1000-029 Lisboa, Portugal
daniel.goncalves@inesc-id.pt
[2] Faculdade de Ciências / University of Lisbon,
Edifício C6, Piso 3. Campo Grande, 1749-016 Lisboa, Portugal
lmc@di.fc.ul.pt
[3] School of Computer Science / University of Manchester,
LF1 Kilburn Building, Oxford Road, M13 9PL, Manchester, UK
markel.vigo@manchester.ac.uk

Abstract. In this document we propose the creation of a Mobile Accessibility Workshop at Interact 2011. Mobile Accessibility is an area that has grown both in importance and number of researchers in recent years. Bringing them together in a workshop would be fruitful and lead to synergies and major developments in the area.

Keywords: Mobile Accessibility, mobile computing, accessibility, workshop.

1 Topic and Objectives

In recent years, we have witnessed an increasing importance of mobile devices. They pervade our daily lives, not only just in the form of feature phones, but also as smartphones and tablets. Tablets in particular have been the object of much attention, lately. They are set to be one of the fastest growing mobile device markets. What is more, they have the computing power of small computers. This is also true for the most advanced smartphones and mobile operating systems. We are, thus, in the cusp of a fundamental change in how we relate to information and others, accessible at all times and places with the help of mobile devices.

Alas, this change might be barred to a large number of people, suffering from a wide range of disabilities. Mobile devices are increasingly visual, making them hard to use by the blind or other vision-impaired people. Keyboards are steadily being replaced by touchscreens, without tactile feedback, introducing additional barriers. Tetraplegic users have limited mobility of their limbs and hardly are able to pick up the devices or use their fingers for precise pointing. Additionally, mobile devices are being used as mediators to reach distal interfaces (e.g. interactive TV, information kiosks, etc.) providing people with disabilities with the means to potentially overcome physical barriers and freely interact with the environment.

Moreover, the mobile aspect often creates situations where those interaction modes are cumbersome or socially unacceptable, even for non-impaired people. Riding a bicycle on a mountain trail and looking at the maps at the same time is difficult and dangerous. Timely responding to urgent and confidential solicitations on a public site

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 734–735, 2011.

may be distressing or embarrassing. These are considered situational disabilities that often introduce similar problems to the use of mobile apparatus. Overall, the list of problems is immense.

Fortunately, there is a nascent area of research, Mobile Accessibility, where researchers are focusing on solving accessibility problems in mobile devices and settings. Bringing together Mobile Computing and Accessibility, it is the source of synergistic works that have the potential of deeply transforming how we look at mobile devices, and to shape the ongoing mobile revolution. Work in this area ranges from new text introduction techniques to multimodal interaction solutions, mobile assistive technology, prototyping tools, navigation tools, theoretical models, etc. A wide range of users and situations is also considered (blind, tetraplegic, elderly, deaf, etc.). As a result, solutions provided to people with disabilities could be applied to the aforementioned situationally impaired users.

Papers on Mobile Accessibility have become an increasing presence in conferences such as CHI, Assets, W4A and MobileHCI, proof of a growing and active community. However, there has been a lack of a single well defined venue for the topmost researchers in the area to meet and exchange ideas and results. We propose, in this first Mobile Accessibility Workshop, to provide such a venue. It is our objective to bring together researcher and practitioners with clear advantages in terms of critical mass, synergies, collaborations and networking, and unmistakable benefits for the area.

2 Participants

Organizers: The workshop will be organized by Prof. Daniel Gonçalves, Prof. Luís Carriço and Markel Vigo, all of them researchers in mobile accessibility, with several papers published in the area.

Intended Audience: Researchers and practitioners of Mobile Accessibility, both from academia and industry, in any of its facets (as described in the introduction).

Expected participants: we expect to attract 12-18 participants.

3 Organization

Duration: the workshop will last an entire day

Format: the workshop will consist, mainly, of peer-reviewed paper presentations by its participants. We'll solicit submissions using the Interact short paper template.

Extended versions of the best papers will be chosen for submission to a Special Issue of the Universal Access in the Information Society Journal

Promoting and Supporting Healthy Living by Design

Gordon Baxter[1], Lisa Dow[1], Stephen Kimani[2], and Nilufar Baghaei[3]

[1] School of Computer Science, University of St Andrews, St Andrews KY16 9SX, UK
{gdb,lisa}@cs.st-andrews.ac.uk
[2] Institute of Computer Science, JKUAT, P.O. Box 62000, 00200, Nairobi, Kenya
stephen.kimani@gmail.com
[3] Dept. Of Computing, Unitec Institute of Technology, Auckland 1142, New Zealand
nbaghaei@unitec.ac.nz

Abstract. Chronic diseases are the main causes of premature deaths, and the number of these deaths keeps growing. People often do not understand, however, that by changing their diet and how much they exercise, they can drastically reduce their risk of being affected by chronic disease. The key to moderating people's behaviour lies in raising awareness of the links between lifestyle and chronic disease and in supporting the adoption and maintenance of a healthy lifestyle. Despite rises in global spending on health care, the pressure on resources is growing as people live longer. With people already using technology for medical information, it is an opportune time to develop technologies that can be used to raise public awareness of the links between lifestyle choices and chronic disease, and facilitate behavioural change.

1 Introduction

Worldwide, chronic diseases such as cardiovascular disease, stroke and cancer are the main causes of premature deaths. By 2030 nearly 75% of deaths will be caused by chronic diseases [1]. The majority of these deaths are the result of a small group of personal lifestyle risk factors such as poor diet, tobacco and alcohol use, and lack of exercise. The incidence of these risk factors keeps growing as people adopt diets high in fats, salt and sugars whilst working longer and exercising less.

From 1997 to 2007, health expenditure grew quicker than national income in most OECD countries [2]. Pressure on the increased resources, however, continues to rise as people live longer. Health care sectors are increasingly unlikely to be able to cope with the rising numbers of people falling ill because of their unhealthy lifestyle.

Many deaths could be prevented if people were aware of the links between lifestyle and chronic diseases, and adopted a healthier lifestyle. People already use technology to increase their knowledge of health issues, e.g., in 2009, in the UK, 14% of people had searched the internet for medical information; a figure expected to rise to 37% by 2020 [3]. It therefore makes sense to develop technologies to raise awareness about links between lifestyle and chronic diseases, and to help people adopt and maintain a healthier lifestyle. In addition to being designed to be usable, these technologies must be: attractive (so people want to use them); effective (in promoting and supporting healthy living); and acceptable (by fitting in with people's everyday lives) [4].

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 736–737, 2011.

2 Goals

This one-day workshop will provide a forum to address current and future issues of designing resources to promote and support healthy living. We want to make learning about, and following a healthy lifestyle enjoyable activities for everybody. This requires an interdisciplinary approach, so we will bring together researchers and practitioners from psychology, health care, computer science, human factors, design and sociology to understand how we can exploit existing technologies (such as gaming [e.g., 5] and Web 2.0 [e.g., 6]), and establish new collaborations to investigate novel ways of promoting and supporting healthy living.

3 Workshop Structure

The workshop will comprise a mixture of invited talks, presentations, and discussion sessions, and there will be a session for demonstrations of existing tools. Participants will submit position papers; accepted papers will be included in the workshop proceedings. The workshop will conclude with a general discussion about future directions to maintain the momentum created by the workshop.

4 Participants

The workshop is aimed at researchers and practitioners from all areas with an interest in preventing chronic disease by promoting and supporting healthy living. These areas include, but are not limited to usability, health and social care, health promotion, ubiquitous computing, health informatics and computer gaming.

References

1. World Health Organization. Obesity and Overweight. Fact sheet Number 311 (2006), `http://tinyurl.com/ys95a7` (accessed 06/05/11)
2. Organization for Economic Co-operation and Development. Health at a Glance: OECD Indicators (2009), `http://www.oecdilibrary.org/oecd/content/serial/19991312` (accessed 06/05/11)
3. Future Foundation. The Future of Remote Diagnosis and Prescription Services: An analysis commissioned by Lloyds Pharmacy (2009), `http://tinyurl.com/nmez9p` (accessed 06/05/11)
4. Steele, M.: An investigation into the usability of health promotion leaflets. Unpublished MSc thesis, University of St Andrews (2009)
5. Berkovsky, S., Coombe, M., Freyne, J., Bhandari, D., Baghaei, N.: Physical activity motivating games: Virtual rewards for real activity. In: Proceedings CHI 2010, pp. 243–252. ACM Press, New York (2010)
6. Baghaei, N., Freyne, J., Kimani, S., Smith, G., Berkovsky, S., Bhandari, D., Colineau, N., Paris, C.: SOFA: An online social network for engaging and motivating families to adopt a healthy lifestyle. In: Proceedings OZCHI 2009, pp. 269–272. ACM Press, New York (2009)

Re-framing HCI through Local and Indigenous Perspectives

Jose Abdelnour-Nocera[1], Masaaki Kurosu[2], Torkil Clemmensen[3], Nic Bidwell[4], Ravi Vatrapu[3], Heike Winschiers-Theophilus[5], Vanessa Evers[6], Rüdiger Heimgärtner[7], and Alvin Yeo[8]

[1] University of West London
[2] The Open University of Japan
[3] Copenhagen Business School
[4] CSIR-Meraka & Nelson Mandela Metropolitan University
[5] Polytechnic of Namibia
[6] University of Amsterdam
[7] Intercultural User Interface Consulting Germany
[8] Universiti Malaysia Sarawak

Jose.Abdelnour-Nocera@uwl.ac.uk
masaakikurosu@spa.nifty.com, tc.inf@cbs.dk,
nic.bidwell@gmail.com, rv.caict@cbs.dk,
heikew@polytechnic.edu.na, evers@uva.nl,
ruediger.heimgaertner@iuic.de, awyeo@acm.org

Abstract. This one-day workshop aims to present different local and indigenous perspectives from all over the world in order to lead into an international dialogue on re-framing concepts and models in HCI/Interaction Design. The target audience is HCI researchers and practitioners who have experience with working with culture and HCI. The expected outcome of the workshop is a) network building among the participants, b) a shortlist of papers that can be basis for a proposal for a special issue of the UAIS journal, and c) identify opportunities to develop a funded network or research proposal.

Keywords: Indigenous HCI, HCI theory and methodology, localization, globalization, cultural usability.

1 Objective and Theme

One of the current challenges for HCI as a discipline is addressing the tensions created between local cultures and the assumptions, priorities and values embedded in the HCI as a discipline. The objective of this workshop is to further our understanding of these issues and lead to practical recommendations for people researching and implementing HCI from an explicit indigenous perspective. Translating local knowledge into valid and useful HCI tools is not a simple problem, but one that requires re-defining and re-negotiating disciplinary boundaries (and connections) and the subject and object of the interaction design. Focusing on local or indigenous awareness and practices in design pushes the envelope in a very exciting way. For instance, the democratic values of equal participation driving user-centred design are not necessarily shared by local communities which prioritize respecting the views of their leaders. Addressing these

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 738–739, 2011.

gaps requires a fresh look at how diverse disciplines and professions explore and conceptualize the relation between users, designers and other stakeholders. While the global HCI community has well-defined conceptual and methodological frameworks, there is little research about how local HCI professionals experience, adapt and implement this knowledge, nor how to locate HCI so that it is locally accountable [1]. To progress this research we must start by better understanding relationships between HCI concepts and methods and their meanings to local and indigenous groups. Universal perspectives on HCI like ethnology and ethnography, e.g., technomethodology [2], and national culture models [3] and activity theory [4] have all had an impact in the design of interactive systems for culturally different users, but the potential contribution of explicitly local or indigenous perspectives, approaches and experiences with HCI, see e.g., [5], have not become so clear and uniform. Furthermore, the idea of what constitutes a useful and usable system in different cultural contexts remains partially explored at the very least. In addition, a simple localization, i.e. the adaptation of products and systems developed in advanced countries to developing countries, may take the form of cultural invasion or sometimes a technological colonization. Thus, we need to support those who are interested in improving the Quality of Life (QOL) of people in developing countries by co-operating the user research and designing something new that will contribute to the QOL of people living there, as well as conducting the conventional localization.

2 Workshop Organization and Duration, Target Audience, and Expected Outcomes

This one-day workshop will address the goals above through three activities: (a) workshop participants will together compile a list of examples of indigenous HCI problems, (b) participants will present short papers on their own current research; (c) group discussion will address specific issues. The target audience is HCI researchers and practitioners who have experience with working with culture and HCI, e.g., participants in IWIPS, ICIC, IFIP TC 13 SIG on Interaction Design and International Development, and more. The expected outcome of the workshop is a) network building among the participants, b) a shortlist of papers that can be basis for a proposal for a special issue of the UAIS journal, and c) identify opportunities to develop a funded network or research proposal.

References

1. Suchman, L.: Located accountabilities in technology production. Scandinavian Journal of Information Systems 14(2), 7 (2002)
2. Button, G., Dourish, P.: Technomethodology: paradoxes and possibilities. In: CHI 1996, pp. 19–26. ACM, New York (1996)
3. Hofstede, G.: Culture's Consequences: Comparing Values, Behaviors, Institutions, and Organizations Across Nations. Sage Publications, Thousand Oaks (2001)
4. Kaptelinin, V., Nardi, B.: Acting with technology: Activity theory and interaction design. MIT Press, Cambridge (2006)
5. Kurosu, M., Kobayashi, T., Yoshitake, R., Takahashi, H., Urokohara, H., Sato, D.: Trend in Usability Research and Activities in Japan. International Journal of Human-Computer Interaction 17(1), 103–124 (2004)

Software Support for User Interface Description Language

Adrien Coyette[1], David Faure[2], Juan González-Calleros[3], and Jean Vanderdonckt[1]

[1] Université catholique de Louvain, Louvain School of Management
Louvain Interaction Laboratory, Place des Doyens 1 - 1348Louvain la Neuve, Belgium
{adrien.coyette,jean.vanderdonckt}@uclouvain.be
[2] Thales Research & Technology France - Campus de Polytechnique
F-91767 Palaiseau Cedex (France)
david.faure@thalesgroup.com
[3] University of Puebla, Faculty of Computer Science
Avenida San claudio y 14 sur, Puebla, Pue. (Mexico)
juan.gonzalez@cs.buap.mx

Abstract. A User Interface Description Language (UIDL) is a formal language used in Human-Computer Interaction (HCI) in order to describe a particular user interface independently of any implementation. Considerable research effort has been devoted to defining various meta-models in order to define rigorously the semantics of a UIDL. These meta-models adhere to the principle of separation of concerns. Any aspect of concern should univocally fall into one of the following meta-models: context of use (user, platform, environment), task, domain, abstract user interface, concrete user interface, usability (including accessibility), workflow, organization, evolution, program, transformation, and mapping. Not all these meta-models should be used concurrently, but may be manipulated during different steps of a user interface development method. In order to support this kind of development method, software is required throughout the user interface development life cycle in order to create, edit, check models that are compliant with these meta-models and to produce user interfaces out of these methods. This workshop is aimed at reviewing the state of the art of software support for a UIDL in the context of any development method (e.g., formal method, model-based, model-driven). From this review, a taxonomy of software support for UIDLs will emerge that will serve for describing, comparing, and exploring software support for UIDLs.

Keywords: Context of use, Model-driven architecture (MDA), Model-driven engineering (MDE), Service Oriented Architecture (SOA), situation engineering, user interface description language (UIDL).

1 Theme, Goals, and Relevance

A User Interface Description Language (UIDL) is a formal language used in Human-Computer Interaction (HCI) in order to describe a particular User Interface (UI) independently of any implementation technology. As such, a UI may involve different interaction modalities (e.g., graphical, vocal, tactile, haptic, multimodal), interaction techniques (e.g., drag and drop) or interaction styles (e.g., direct manipulation, form filling, virtual reality). A common fundamental assumption of most UIDLs is that UIs

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 740–741, 2011.

are modeled as algebraic or model-theoretic structures that include a collection of sets of interaction objects together with behaviors over those sets. Significant examples of UIDLs include: UIML (www.uiml.org) [5], useML (http://www.uni-kl.de/pak/useML/), MariaXML, UsiXML (www.usixml.org), and XIML (www.ximl.org). Various UIDLs have been subject to discussion, understanding their common ground and their subsumed approach, comparative analysis, and their consideration for standard. Sometimes, alternative approaches have been considered and compared within a same UIDL such as UsiXML.A UIDL can be used during:

- *Requirements analysis*: in order to gather and elicit requirements.
- *Systems analysis*: in order to express specifications those address the aforementioned requirements.
- *System design*: in order to refine specifications depending on the context of use.
- *Run-time*: in order to realize a UI via a rendering engine.

The design process for a UIDL encompasses defining the following artefacts:

- *Semantics*. They express the context, meaning and intention of each abstraction captured by the underlying meta-models on which the UIDL is based on. Meta-Models are normally represented by means of UML Class Diagrams, OWL or other conceptual schemas. Semantics are usually conveyed using natural language.
- *Abstract Syntax*. It is a syntax that makes it possible to define UI models (in accordance with the UIDL semantics) independently of any formalism.
- *Concrete Syntax/es*. They are (one or more) concrete representation formalisms intended to express syntactically UI Models. Many UIDLs has an XML-based concrete syntax. In fact XML has been proven to be extremely useful in describing UIs according to the different levels of the Cameleon Reference Framework (CRF) [1] and for adapting UIs according to adaptation dimensions of the Similar Adaptation Space (SAS) [2].
- *Stylistics*. They are graphical and textual representations of the UIDL abstractions that maximize their representativity and meaningfulness in order to facilitate understanding and communication among different people. Stylistics are typically used by models editors and authoring tools.

Many UIDLs reveal themselves as a markup language that renders and describes graphical user interfaces and controls. But a UIDL is not necessarily a markup language (albeit most UIDLs are) and does not necessarily describe a graphical user interface (albeit most UIDLs abstract only graphical user interfaces). Figure 1 shows a general software architecture depicting typical software support for a UIDL. The workshop is aimed at defining a taxonomy for such a software support so that it can be used widely to refer to the same base. It is expected to review existing software support in the light of this taxonomy.

References

1. Calvary, G., Coutaz, J., Thevenin, D., Limbourg, Q., Bouillon, L., Vanderdonckt, J.: A Unifying Reference Framework for Multi-Target User Interfaces. Interacting with Computers 15(3), 289–308 (2003)
2. Vanderdonckt, J., Coutaz, J., Calvary, G., Stanciulescu, A.: Multimodality for Plastic User Interfaces: Models, Methods, and Principles. In: Tzovaras, D. (ed.) Multimodal User Interfaces: Signals and Communication Technology. LNEE, pp. 61–84. Springer, Heidelberg (2007)
3. Workshop web site at: `http://itea.defimedia.be/node/130`

User Experience in Cars

Manfred Tscheligi[1], Albrecht Schmidt[2], David Wilfinger[1], Alexander Meschtscherjakov[1], and Andrew L. Kun[3]

[1] ICT&S Center, University of Salzburg, 5020 Salzburg, Austria
{manfred.tscheligi,david.wilfinger,
alexander.meschtscherjakov}@sbg.ac.at
[2] University of Stuttgart, 70569 Stuttgart, Germany
albrecht.schmidt@acm.org
[3] University of New Hampshire, Electrical and Computer Engineering Department, Kingsbury Hall, Durham, NH 03824, USA
andrew.kun@unh.edu

Abstract. This workshop will address two emerging fields within the HCI community: user experience (UX) and the automotive context. It will bring HCI experts together to discuss UX factors for the specific characteristics of car interiors and automotive user interfaces. It targets the development of a better view of UX within the whole car (driver, front seat, backseat area) beyond traditional marketing instruments known within the automotive industry.

1 Topic, Objectives and Outcomes

HCI is always dependent on the context where the interaction happens and therefore contingent on a multitude of factors. Especially the automotive context has its own and very specific characteristics. Car cockpits have become more complex due to a variety of assistance systems and infotainment capabilities. In addition cars are often a very emotional topic for humans. This workshop will explore the inclusion of user experience (UX) aspects in researching and developing automotive user interfaces.

With cars becoming interactive systems that offer besides transportation all kinds of information access, assistance, as well as entertainment, UX research has become a major challenge in this context. Alternative interaction means for example will be offered for the driver as well as for front and rear seat passengers (e.g. supporting transparency related to the driving situations or supporting co-pilot activities). The introduction of such interactive systems will require a deep understanding of UX in the car (i.e. car as a very special context) in order to be successful. For that purpose a common understanding of user experience and its effects on car technology (and all different subparts) usage is needed, as well as an in depth discussion on the applicability of UX methods that were established in other application contexts. This workshop therefore addresses the following issues:

- Understanding of UX in the automotive context in a holistic way
- Looking into different sub-domains: the driver, the front seat, the backseat area, interconnection to the outside world
- Researching the relevance of traditional UX factors in the automotive context and identification of new UX factors especially relevant for the car

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 742–743, 2011.

- Contextual analysis and evaluation methods for UX in the automotive context
- Discussion of car laboratory setups in comparison to in-situ methods for car UX research
- Automotive user interface frameworks and toolkits
- New concepts for in-car user interfaces increasing UX and supporting safety
- Experience design in the car
- Looking into the future: the ubiquitous car in a mobile society

The intended outcomes of this workshop include the creation of a common understanding and perspective of the challenges related to UX and automotive interfaces. For that purpose the workshop will discuss perspectives on how to address these challenges in future research and development. After the workshop accepted papers will be made available on the workshop website. The publication of a revised version of papers as a special issue of a journal (such as Personal and Ubiquitous Computing or Interacting with computers) will be considered and an Edited Book is envisaged (see more details on http://workshops.icts.sbg.ac.at/interact2011).

2 Workshop Organization

The workshop is intended for: HCI researcher in general, who are interested specifically in the automotive context; UX experts, who are willing to adapt UX models and extend UX factors for the specific characteristics of interaction in the car; automotive user interface designers and engineers both from a scientific, as well as from an industrial perspective. Workshop candidates are requested to send a position paper about their research and link to the workshop theme to the organizers (email: interact2011@hciunit.org).

This is a one-day workshop with breakout sessions, alternated with a moderated group discussion. The workshop will start with an introduction to the workshop topic, followed by introductory presentations. After a break the organizers present the common themes of the submitted papers, grouping them into different sessions. The different groups will then discuss their topics during a first breakout session. After the lunch break a short alignment of pre-lunch discussions will happen in the whole group. This continues into a second round of breakout sessions to further discuss grouped topics. After a coffee break the results of the breakout sessions will be discussed in the whole group. The conclusions of the workshop will be worked out and follow up activities will be specified.

Organizers' backgrounds: Manfred Tscheligi is professor for HCI & Usability at the University of Salzburg. He is currently acting as Conference Chair for the 3rd Conference AutomotiveUI 2011. Albrecht Schmidt is a professor for Human-Systems Interaction and Cognitive Systems at the University of Stuttgart and was Conference Co-Chair of AutomotiveUI 2010 and 2009. Andrew Kun is Associate Professor of Electrical and Computer Engineering at the University of New Hampshire and was Program Co-Chair of AutomotiveUI 2010. Alexander Meschtscherjakov and David Wilfinger are Research Fellows at the HCI Unit of the ICT&S Center (University of Salzburg) working in the area of automotive user interfaces.

User Interaction Techniques for Future Lighting Systems

Dzmitry Aliakseyeu[1], Jon Mason[1], Bernt Meerbeek[1], Harm van Essen[2], Serge Offermans[2], and Andrés Lucero[3]

[1] Philips Research Europe, 5656 Eindhoven, The Netherlands
[2] TU Eindhoven, Industrial Design department, Eindhoven, The Netherlands
[3] Nokia Research Center, Visiokatu 1, 33720 Tampere, Finland
{dzmitry.alisksееyeu,jon.mason,bernt.meerbeek}@philips.com
{h.a.v.essen,s.a.m.offermans}@tue.nl, andres.lucero@nokia.com

Abstract. LED-based lighting systems have introduced radically new possibilities in the area of artificial lighting. Being physically small the LED can be positioned or embedded into luminaires, materials and even the very fabric of a building or environment. Together with new functionality and flexibility comes complexity; the simple light switch is not anymore sufficient to control our light. This workshop explores new ways of interacting with light. The goal is to define directions for new forms of user interaction that will be able to support the emerging LED-based lighting systems.

Keywords: Lighting, User Interaction, LED, Smart lighting.

1 Introduction

The Light Emitting Diode (LED) has caused a profound change within the lighting industry. This is due in part to the LED's key properties of being physically small, highly efficient, digitally controlled and soon, very cheap to manufacture. Being physically small the LED can be positioned or embedded into luminaires, materials and even the very fabric of a building or environment [1]. The price to pay for all this functionality and flexibility is complexity. In the past, the single light bulb was controlled using a single switch; on and off. LED-based lighting systems can easily consist of hundreds separate light sources, with each source having many individually controllable parameters including colour, intensity, and saturation. With this high complexity, end-users cannot be expected to fully control all aspects of the lighting system. One direction that is being explored is to enrich lighting systems with sensor networks that will enable automatic lighting control that is based on contextual information [2]. However in many situations, such as setting up an atmospheric light, an explicit user interaction will still be required. Moreover, as functionality and complexity of light systems grow, the mapping between the sensors data and the desired light outcome will become fuzzy and will require an explicit user interaction for fine tuning the outcome or for adjusting the mapping between sensor input and light output. Thirdly, explicit interaction can be desired to allow users to feel in control while interacting with intelligent lighting systems. The light switch therefore in many situations will need to be replaced by novel forms of interactions that offer

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 744–745, 2011.

richer interaction possibilities such as tangible, multi-touch, or gesture-based user interfaces. As proliferation of LED light continues, it becomes more important to go beyond scattered design efforts [2, 3] and systematically study user interaction with emerging lighting systems. The goal of this workshop is to take the first steps in this direction.

2 Goals of the Workshop

The focus of this workshop is on formulating key research challenges for user interaction with future lighting systems, creating initial design guidelines, and proposing novel interaction techniques for these systems. The goals of the workshop:

1. Make a first step toward expanding the design space of interactive technologies to include new forms of decorative, ambient, and task lighting.
2. Identify key challenges of UI for controlling new forms of lighting systems.
3. Establish a link with existing interaction paradigms that can be (re-)used for control of future lighting systems.

During the workshop, we would like to address and discuss the following questions:

- What design opportunities for interactive technology exist in the context of the new forms of lighting?
- What forms and types of (existing) interaction are suited for emerging lighting systems (in particular tangible, gesture and multi-display interaction techniques)?
- What forms of interaction are best suited for a global control (e.g. atmosphere) and what for a point control (e.g. task lighting)?
- How to balance between explicit user control and internal system control?
- How to use a lighting infrastructure as ambient displays and how to combine it with its primary function i.e. illumination?
- What is the impact of the proposed interaction techniques for complex lighting systems in other domains? What is the generalizability of these techniques?

To address the workshop questions we are inviting researchers to submit position papers that discuss or present new forms of interaction techniques for lighting. This topic deals with the research and design of new forms of user interaction or adaptation of existing ones to emerging lighting systems. The topic should attract researchers and designers working on new forms of interaction, LED lighting, smart lighting systems and who are interested in exploring UI for new emerging types and forms of luminaires and lighting systems.

References

1. Price, C.: Light Fantastic. Digital Home Magazine (November 2003)
2. Bhardwaj, S., Ozcelebi, T., Lukkien, J.: Smart lighting using LED luminaries. In: Proc. of PERCOM Workshops, pp. 654–659. IEEE, Los Alamitos (2010)
3. Lucero, A., Lashina, T., Terken, J.: Reducing Complexity of Interaction with Advanced Bathroom Lighting at Home. I-COM, Oldenbourg 5(1), 34–40 (2006)

Values in Design - Building Bridges between RE, HCI and Ethics*

Christian Detweiler[1], Alina Pommeranz[1],
Jeroen v.d. Hoven[1], and Helen Nissenbaum[2]

[1] Delft University of Technology, The Netherlands
{c.a.detweiler,a.pommeranz,m.j.vandenhoven}@tudelft.nl
[2] New York University, USA
helen.nissenbaum@nyu.edu

1 Objectives

Designing for values has become increasingly important for technology development. In many technological systems (medical applications, social networks etc.) values (privacy, autonomy, trust etc.) play a role and are sometimes violated. In working with stakeholder requirements or user needs, various design methods in requirements engineering (RE) [3] and human computer interaction (HCI), in specific user-centered (UCD), deal with "soft issues" [4], "social issues", "people issues" or values. At the same time, applied ethics has begun to pay attention to design. We believe that many of the approaches could complement each other in useful ways. The aim of this workshop is to bring together people from different disciplines to share knowledge and insights about how to account for values in technology design, and to work towards integrating approaches, thereby putting value conscious design approaches (e.g. values-in-design [1] or value sensitive design [2]) to practice.

2 Theme and Topics

The main theme is the interdisciplinary exchange of knowledge, experiences and new ideas on *values in technology design*. The following list reflects possible topics:

- Designing for specific values, experiences from value-oriented projects
- Understanding and relating different notions of values
- Integrating value-oriented methods with software engineering/ design methods
- Accounting for values in experience oriented designs, e.g. "fun" as a value
- Values used in evaluations of User Experience (UX)
- Dealing with soft issues, social issues, and people issues in RE, Value-based RE
- Value elicitation, dealing with a variety of stakeholder values
- Capturing and reusing value knowledge (scenarios, design patterns, etc.)
- Values in Persuasive Technology, Ambient Intelligence and Personal Informatics
- Values in specific domains (health care, military, crisis management)

* All authors of this proposal are to be considered key organizers of the workshop.

P. Campos et al. (Eds.): INTERACT 2011, Part IV, LNCS 6949, pp. 746–747, 2011.

The target audience includes HCI and social science researchers interested in social aspects of technology and values in design. We also aim at researchers focusing on soft issues in RE, UCD and moral philosophers interested in ethics and technology. For interactions between academia and industry we will invite (interaction) designers and software engineers/developers.

3 Workshop Format

This is a one-day workshop divided into two parts. In the first half we will have four consecutive panel sessions: the values turn in design - “soft issues” in RE and UCD; the design turn in applied ethics; values across disciplines; and values in industry. Panels will be led by representatives in the respective field. Each panel will be initiated by short presentations by authors of accepted papers followed by a discussion with all workshop participants. In the second part, we will create groups of 4-5 participants with different backgrounds (RE, HCI and Ethics). Each group will be given a design case to work on using tools and methods proposed by the participants. The aim will be to create an overview of people issues, values, soft issues and (long term) social aspects emerging from each design case. In a debriefing session we will share experiences (regarding the cases and tools used) from the group work with all participants. Participants who wish to present their work as posters or demos will be given the opportunity to do so during the breaks.

4 Expected Outcomes

We expect that the interdisciplinary collaboration in the workshop will lead to new ideas and more practical approaches to the inclusion of values in technology design. More specifically, an outcome of the workshop will be the publication of a volume of extended versions of work submitted to the workshop highlighting the state of the art in addressing values in design. Additionally, insights from the panel discussions and interdisciplinary work will form the basis for a number of publications in a special issue of the “Ethics and Information Technology” journal focusing combining existing approaches, and on new approaches that emerge from the workshop.

References

1. Flanagan, M., Howe, D., Nissenbaum, H.: Embodying Values in Technology: Theory and Practice. In: van den Hoven, J., Weckert, J. (eds.) Information Technology and Moral Philosophy. Cambridge University Press, Cambridge (2008)
2. Friedman, B., Kahn, P., Borning, A.: Value Sensitive Design and Information Systems. In: Zhang, P., Galletta, D. (eds.) Human-Computer Interaction and Management Information Systems: Foundations, pp. 348–372. M.E. Sharpe, New York (2006)
3. Proynova, R., Paech, B., Wicht, A., Wetter, T.: Use of personal values in requirements engineering – A research preview. In: Wieringa, R., Persson, A. (eds.) REFSQ 2010. LNCS, vol. 6182, pp. 17–22. Springer, Heidelberg (2010)
4. Thew, S., Sutcliffe, A.: Investigating the Role of ’Soft Issues’ in the RE Process. In: 16th IEEE International Requirements Engineering, RE 2008, September 8-12, pp. 63–66 (2008)

Author Index